eTheses
maximising our postgraduate potential

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Ellis, Lucy Anne. 2000. An instrumental study of alveolar to velar assimilation in slow and fast speech using EPG and EMA techniques. PhD thesis. Queen Margaret University.

Accessed from:

http://etheses.qmu.ac.uk/171/

Repository Use Policy

The full-text may be used and/or reproduced, and given to third parties for personal research or study, educational or not-for-profit purposes providing that:

- The full-text is not changed in any way
- A full bibliographic reference is made
- A hyperlink is given to the original metadata page in eResearch

eResearch policies on access and re-use can be viewed on our Policies page:
http://eresearch.qmu.ac.uk/policies.html

http://etheses.qmu.ac.uk

An instrumental study of alveolar to velar assimilation in slow and fast speech using EPG and EMA techniques

Lucy Anne Ellis

PhD
Queen Margaret University College
2000

An instrumental study of alveolar to velar assimilation in slow and fast speech using EPG and EMA techniques

ABSTRACT

This thesis evaluates the widely-held notion that place assimilation is (i) more frequent at faster rates of speech and (ii) a gradual phonetic process. The latter view is based on previous small-scale EPG studies which showed evidence of partial alveolar assimilations lacking complete stop closure on the alveolar ridge but with a residual tongue body gesture. For the present study, EPG data from 10 speakers were collected. Two experimental sequences, /n#k/ and /ŋ#k/, embedded in meaningful sentences, were produced by subjects 10 times each in a slow/careful style and 10 times each in a fast/casual style. The first sequence captures the potential site of assimilation and the second is a lexical velar-velar sequence with which cases of complete assimilation can be compared. The results showed that, overall, assimilation was more frequent in fast speech than in careful speech, although timing analysis revealed that assimilation is not the automatic consequence of rate-induced changes in intergestural timing of /n#k/. In fast speech, six of the ten speakers showed relatively consistent assimilatory preferences: they either produced only complete assimilations or they never assimilated. However, four speakers showed considerable intra-speaker variability. Two of the four produced either full alveolars or complete assimilations in the manner of a categorical opposition (complete assimilations were indistinguishable from control /ŋ#k/ sequences). The other two speakers produced a continuum of forms that could be ranked from full alveolars to complete assimilations via partial assimilations. Using the same stimuli, a follow-up combined EPG/EMA study was carried out, the purpose of which was to look for reduced coronal gestures undetectable in tongue-palate contact-only data. Two 'categorical' assimilators were re-recorded and these gestures were not found. This supports the interpretation that for some speakers assimilation is determined at a higher level through the application of a cognitive rule, while for others variation is 'computed on-line' during speech production itself. Current phonological models of assimilation are found to be unable to capture both gradient effects and more radical feature-sized substitutions under a single framework.

Declaration

I declare that the composition of this thesis is my own and that the work contained herein is my own unless otherwise indicated.

Lucy Anne Ellis

29^{th} May 2000

Acknowledgments

I would like to take this opportunity to thank the following people for their contribution to this thesis:

Janet Beck and Nigel Hewlett (supervisory advice); Peter Roach (judicious comments); James Scobbie (many valuable discussions and enlivening feedback on an earlier draft); Alan Wrench (technical expertise); Steve Cowen (sound advice on software and presentational problems too numerous to mention); my experimental subjects; officemates Fay Windsor, Anne Walton, Ben Matthews and Suzanne Fuchs (fun and friendship); Heather McLeish (administrative support) and finally I would like to extend my thanks to my Director of Studies, Bill Hardcastle, for his never-failing interest, skill and friendship.

Owen, you have given me all the support I needed over the years to see me through this and for that I am so grateful.

I dedicate this work to Mum and Dad who have always been there for me.

TABLE OF CONTENTS

(1) Ellis, L. & Hardcastle W. J. (1998) An EPG study of alveolar to velar coarticulation in fast and careful speech: some preliminary observations. In Artemis Alexiadou, Nanna Fuhrkop, Ursula Kleinhenz & Paul Law (eds.) *ZAS Papers in Linguistics*, 11, 105-120

(2) Ellis, L. & Hardcastle W. J. (1999) An instrumental study of alveolar to velar assimilation in careful and fast speech. *Proceedings of the XIVth International Congress of Phonetic Sciences,* Volume 3: 2425-2428

List of Figures

List of Tables

CHAPTER ONE

Review of the Literature

1.0 INTRODUCTION

In a general review of approaches to assimilation phenomena in English, Local (1992: 206) comments:

> They are remarkably uniform not only in what they say but in the ways they say it. Despite some differences in formulation and in representation, they tell the same old story: when particular things come into contact with each other one of those things changes into, or accommodates in shape to, the other…*What* are these things? *Where* are these things?

A major concern for instrumental research into the articulatory details of place assimilation in recent decades has indeed been the 'where?', and to a lesser extent, the 'what?' of assimilation. This concern has been given explicit expression more recently, in the context of work that asks how higher-order grammatical rules governing assimilation interact with low-level processes such as physical properties of the vocal mechanism. The study reported in this dissertation addresses this question.

As well as a review of particular articulatory, acoustic and perceptual studies that have a bearing on this issue, with a focus on alveolar to velar assimilation, this introductory chapter will provide some scene-setting basic principles of assimilation. These include: definition and description of assimilation; the concept of coarticulation and how it relates to assimilation; phonological modelling; factors that are known to affect assimilation; the relationship to models of speech production and psycholinguistic processing. Following this introduction, a series of research questions will be posed which have arisen partly through gaps in knowledge about assimilation, the methodological shortcomings of previous, closely related studies and finally, the need for clarification of certain trends in assimilation.

1.1 CLARIFICATION OF TERMS

This section presents some views on the concepts *coarticulation* and *assimilation* and sets out the reasons behind the use of the term assimilation in this dissertation to describe the changes that an alveolar may undergo when it appears before a word-initial velar plosive.

There has been a tendency in recent years to conflate assimilation and coarticulation into a single concept and sometimes they are even used synonymously. Traditionally, however, the former is used to describe a phenomenon which is extensive enough to be detected auditorally whereas the latter is used to interpret phenomena only adequately detected by instrumental techniques, acoustic or articulatory. Jones (1932) introduces a separation of 'similitude' and 'assimilation'. Similitude describes a situation where two sounds influence each other so that they become alike but not to the extent that they radically change their phonemic identity, for example when the usually voiced /l/ becomes partly devoiced in 'flame' [fl̥eɪm]. Jones uses assimilation to describe a situation where phonemic identity is affected and one phoneme is thought to be replaced with that of an adjacent one, as in 'spaceship' [speɪʃ(ʃ)ɪp]. Barry (1985: 3) distinguishes between 'those cases where the result of the assimilation is an allophone of the phoneme undergoing assimilation (co-articulation), and those cases where the result of an assimilation belongs to a different phoneme' and Hawkins (1984) proposes that not only are co-articulations allophonic and assimilations phonemic, but that the former are automatic and exceptionless, the latter often optional with exceptions. Lindblom (1983) similarly refers to assimilations as categorical.

Both assimilation and coarticulation can be expressed in terms of directionality, however. Coarticulation can be anticipatory or perseverative and some segments in English can be assimilated backwards or forwards. On the subject of direction, Daniloff and Hammarberg (1973) consider that carryover (left-to-right) coarticulation is always due to inertial effects on the speech organs while anticipatory coarticulation is always a 'deliberate' phonological process inducing categorical changes in the description of segments undergoing changes. The assimilation of a word-final alveolar to a following word-initial velar falls into the latter category.

Laver (1994:152) deals with assimilation and coarticulation in the context of 'phonetic setting' which is defined as:

> any co-ordinatory tendency underlying the production of the chain of segments in speech towards maintaining a particular configuration or state of the vocal apparatus. More specifically, a setting consists of one or more featural properties held in common by two or more speech segments in close proximity in the stream of speech. The segments concerned can be regarded as carriers of the setting.

Within this definition, assimilation and coarticulation are distinguished by means of how many segments a setting spans. 'In assimilation, the minimum span of a setting extends over two segments, located on each side of a word-boundary. In the case of co-articulation, we have seen that the span of a setting in the chain of speech can range from a minimum of two adjacent segments up to six or seven segments.' (p.152) Indeed, Daniloff and Moll (1968) found that lip protrusion for /u/ can be detected at the start of CCCCV sequences. Further on up Laver's hierarchy of span, is segmental harmony (which are said to cover all the segments within a word, or sometimes within a given morphological part of a word) followed by paralinguistic communication where the span can extend as far as over a whole utterance pronounced in a particular tone of voice and culminating in the extralinguistic domain where every single utterance produced by a particular speaker is phonetically coloured to some degree by his or her personal quality.

Clark and Yallop (1990) do not separate assimilation from coarticulation as different processes. Their discussion of assimilation is introduced as a departure from coarticulation by virtue of it having 'a longer tradition'. They do go on, however, to point out that there is a common understanding of assimilation which revolves around 'ill-defined conventions about the nature of transcription.'(p.121) Clark and Yallop consider assimilation to be a consequence of coarticulatory effects and because of this causal relationship they believe that the term assimilation is more widely used than coarticulation especially in terms of consonants.

By far the most influential definition of coarticulation is that found within standard generative phonology. Chomsky and Halle's *Sound Pattern of English* (1968) placed assimilation and coarticulation in opposition to each other. Coarticulation pertains to the 'Performance' domain of speech and is said to be brought about by universal physical properties of the speech apparatus. Assimilation on the other hand is governed by language-specific phonological rules (which operate by the modification of features) and is thus part of the grammar or 'Competence' domain of speech. Starting from this framework, experiments to determine whether variability in speech stems from assimilation or coarticulation assess whether these effects are universal (Performance) or language-specific (Competence) and whether the type of modification is categorical and thus 'intentional' or is just background 'articulatory accommodation'.

A radical and relatively recent re-formulation of the concepts assimilation and coarticulation has taken place, indirectly, with the theory of Articulatory Phonology (Browman and Goldstein, 1990; 1992). This theory accounts for all variability phenomena with a single unifying principle. Units of speech production in Articulatory Phonology are gestures which are temporally specified on a 'gestural score'. Both coarticulatory effects and assimilations are explained primarily as increases in gestural overlap (with a possible role for decreases in gestural displacement), assimilation being different to coarticulatory effects only in terms of degree. When speech is rapid and/or casual, adjacent consonantal gestures on separate tiers can overlap so much as to give rise to 'hiding' which is said to correspond to the *perception* of a deleted gesture in the production of, for instance, an alveolar to velar sequence. Since assimilation is said to arise 'naturally', phonological assimilation rules have no part to play. An expanded description of Articulatory Phonology is supplied in Section 1.2.2 below.

This dissertation makes deliberate use of the term assimilation in preference to the term coarticulation. This is because of a number of factors. Alveolars are a special class of speech sounds since they are 'unstable' in a way that other speech sounds are not in English (Gimson, 1960). Some have attributed this instability to the idea that they are phonologically 'underspecified' (see section 1.2.1.1). For this reason, the variability they display should be treated differently from the variability that stems from more continuous and pervasive coarticulatory effects. Secondly, because place assimilation is optional, assimilatory variants are likely to contrast between themselves more strongly and thus be perceived auditorily, unlike coarticulatory variants. However, evidence reviewed below in this chapter suggests that alveolar to velar realisations owe more to gradual coarticulatory effects than discrete phonological effects. For this reason reference will be made to some studies of coarticulation throughout this chapter. Categorical phenomena do, however, exist as will be demonstrated by the results of this present study. The 'received wisdom' about assimilation and coarticulation and what they each describe, particularly with regard to the extent of language-specific effects, requires more empirical investigation. The following sections describe coarticulation and assimilation in more detail, respectively.

1.1.1 Coarticulation

In articulatory terms, coarticulation describes the way in which several different articulators dynamically interact in time and space during the production of successive phonetic segments. A cross-section of the vocal tract at any particular moment during connected speech will reveal a complex configuration associated with more than one segment. An example of coarticulation is the variation that can occur in the production of English /k/. When it appears following a front vowel as it does in the word /k̟iː/ *key*, it is produced for most speakers with the tongue fairly far forward on the palate, but tongue placement will be farther back when /k/ occurs before a back vowel in the word /k̠ɔːt/ *caught*. Coarticulatory effects can also be noted for lip position. In anticipation of the rounded vowel /ɔː/, rounding will be in evidence from the onset of velar closure in *caught* contrasting with the relatively spread state of the lips for velar closure before /iː/ in *key*.

Although the concept of coarticulation describes the continuous and variable activity of speech production, invariant and discrete units are the fundamental structures that underlie this activity. If these invariant and pre-articulatory units were not posited at some level, then for every conceivable sequence of speech units there would have to be a maximally detailed articulatory plan, clearly a huge burden in terms of psycholinguistic computation, storage and access. So, the notion of phoneme-sized units provides a sufficient abstraction from the physical execution of speech to 'encode' it efficiently and meaningfully. On this subject Kühnert and Nolan (1999:8) have commented:

> Ironically, studies of coarticulation itself, based on the premise of phoneme-sized segments at some level of representation, lend independent support to the premise. For instance such studies have conspicuously not shown, to take a hypothetical example, that the onset of lip-rounding is consistently different between the words *caw* and *caught*, or that the degree of velar fronting is consistently different in each of the words *key, Keith,* and *keen*. If each word were represented holistically and independently, this result would `have to be put down to coincidence. Repeated across the lexicon the coincidence would be huge and extraordinary. On the other hand this kind of regularity across words is predicted by a view in which these sets of words are represented with common sequences of abstract elements, for instance /k/, /ɔː/ in the case of the first set and /k/, /iː/ in the case of the second, and in which the articulatory realisation of those sequences is governed by regular principles of integration - that is, by principles of coarticulation.

Another characteristic of coarticulatory effects is that they can be essential information-bearers. The fact that properties of a segment can spread considerably beyond its own domain means that information relating to segments has more chance of being perceived by the listener than if all cues were limited to a segment's own boundaries. Ali et al (1971) discovered that when subjects were presented with only the CV or CVV of a CVC or CVVC sequence (the final consonant and the transition into it were removed) they were still able to predict whether the final C was nasal or non-nasal. This was interpreted as evidence that anticipatory coarticulation can aid the perceptual process. This case has been put more strongly by Maeda (1999) whose study of labialisation during /k/ followed by a rounded vowel showed that lip rounding is in fact perceptually *required*. Results of a perceptual experiment confirmed this and he also found that variability of the rounded lip position during /k/ when followed by /u/ is small whereas before other vowels, lip rounding in /k/ is much more variable.

1.1.2 Assimilation

Standard English phonetics and phonology textbooks usually locate their discussion of assimilation within a general overview of connected speech and connected speech processes (hereafter CSPs). Typically, once the identity of the building blocks of speech have been established (the phoneme), the higher-order architecture of connected speech is described along with possible variations on citation forms such as assimilation, segmental deletion and prosodic restructuring. Sometimes, however, assimilation is approached differently, for instance within the context of coarticulation and speaker-specific variability (Clark and Yallop, 1990). And some discuss assimilation with a view to the practical requirements of the foreign language learner (Gimson, 1989; Roach, 1983). Gimson (1989) begins with a discussion of the abstract nature of the word. If, he says, the word is to be admitted as an abstract linguistic unit, it must be remembered that differences exist between its (often artificial) form in isolation and its form when subject to 'the pressures of its phonetic environment' or effects of different accentual or rhythmic groups of which it is a part. He gives the name 'special context forms' to those forms which typify connected speech. Variations are divided between those that affect the word as a whole, like weak forms in an unaccented context or accentual patterns within larger rhythmic patterns of a complete utterance, and those which affect sounds at word boundaries. These latter variations include junctural assimilations, elision and 'liaison forms' e.g. linking /r/. While explaining the occurrence of variation in place of articulation (i.e. as it occurs in assimilation) during connected speech he notes 'though such changes are normal in colloquial speech, native speakers are usually unaware that they are made' (p.288). We are left to make the assumption that non-native speakers are aware of these variations. Roach (1983) again treats connected speech as something which, for foreign learners, alters what they know about the pronunciation of words in isolation. He remarks: 'It would not be practical or useful to teach all learners of English to produce assimilations; practice in making elisions is more useful, and it is clearly valuable to do exercises related to rhythm and linking.' (p.111) Gimson (1960) has noted in 'Le maître Phonetique' that assimilations of the type *ran quickly* [raŋ kwɪkli], where the burden of meaning rests with the context, do not pose a threat to intelligibility for English-speaking listeners. Other types of assimilatory changes are very much noticed however: 'others will be noticed and characterised by some such term as "slipshod" or "uneducated"' (p.8). He gives the example of the London dialect version of *down the road* [daun nə rəud] which, he says, is considered by RP speakers to be a vulgarism. He also comments that the most commonly assimilated items, alveolars, are the ones most likely to pass unnoticed.

In Gimson's framework (1989) assimilation of word final alveolars comes under 'variations at word boundaries' thus:

/t/ assimilates to /p/ before /p, b, m/ e.g. *that man* and to /k/ before /k, g/ e.g. *that girl.*
/d/ assimilates to /b/ before /p, b, m/ e.g. *good man* and to /g/ before /k, g/ e.g. *good concert*
/n/ assimilates to /m/ before /p, b, m/ e.g. *ten players* and to /ŋ/ before /k, g/ e.g. *ten cups* [1]
/s/ assimilates to /ʃ/ before /ʃ, j/ e.g. *this shop, this year*
/z/ assimilates to /ʒ/ before /ʃ, j/ or /ʃ/before /ʃ/ e.g., *has she?*/hæʒ ʃi/ or /hæʃ ʃi/[2]
(Gimson, 1989: 290)

Gimson notes that alveolars have a relatively high frequency of word final occurrence, especially when inflexional but that the sense of an utterance can be determined by the context, illustrated by examples such as / 'wɒtʃ jə weɪt/ *what's* or *watch your weight*

---

[1] Gimson notes that as a result of word final assimilations /ŋ/ *may* be preceded by vowels other than /ɪ, e, æ, ɒ, ʌ/ a situation which violates the rule that long vowels never appear before /ŋ/ e.g. *I've been* / 'biŋ/ *gardening, She'll soon* / 'suŋ/ *come, his own* / 'əʊŋ/ *car* etc

[2] Gimson notes that final /ð/ may assimilate to /z/ or /s/ before /s, z/ e.g. *I loathe singing* /ai 'ləʊz sɪŋɪŋ/

On the subject of the progression from formal to informal style, Laver (1994) comments: 'the phonemic structure of individual words is often re-organised to reduce the complexity and number of syllables.' (p.67). For example, the maximally formal RP pronunciation of the word 'actually' /aktjʊəlɪ/ can be reduced to the maximally informal pronunciation [aʃlɪ]. Some of these more simple forms occur in quite formal speech. He also notes that in situations where segments are deleted, syllable structures may in fact become more rather than less articulatorily complex such as in casual pronunciations of 'potato' or 'tonight': [pteɪtə], [tnaɪt]. Therefore Laver considers that the relatively complex sequences [pt-,tn-] can in fact be more characteristic of informal than formal speech.

1.2 APPROACHES TO ASSIMILATION

The following sections describe various models and ideas which claim to describe or account for assimilation. Firstly, phonological approaches will be described: segmental phonological formalism and feature geometry formalism. Following this the concept of phonological underspecification is described. Then, some models of speech production are outlined which have an important bearing on phenomena such as assimilation.

1.2.1 Phonological approaches

It is almost axiomatically considered that place assimilation is a discrete process of phoneme substitution brought about by a phonological rule (Spencer, 1996). One form of conventional phonological representation of assimilation, in the case of alveolar to velar assimilation, is of the type:

n → ŋ / ____ {k, g}

Figure 1.1: linear phonological representation of alveolar to velar place assimilation

This purely segmental model treats assimilations as a substitution of a complete segment. A similar model is based on features instead of segments and an example, following Chomsky and Halle's feature set (1968), is given below:

(1)

$$\begin{bmatrix} +\text{coronal} \\ +\text{anterior} \\ -\text{continuant} \end{bmatrix} \longrightarrow \begin{bmatrix} \alpha\ \text{coronal} \\ \beta\ \text{anterior} \\ \gamma\ \text{distr} \end{bmatrix} \Big/ \underline{\qquad} \begin{bmatrix} \alpha\ \text{coronal} \\ \beta\ \text{anterior} \\ \gamma\ \text{distr} \end{bmatrix}$$

Figure 1.2 linear feature phonological model of place assimilation

More recently, autosegmental frameworks have claimed to capture the fact, lacking in the linear models shown above, that assimilations involve certain groups of features being modified as a unit, i.e. features of place of articulation. The organising principle is that features are represented as hierarchically ranked in terms of functional groups. Each level, or tier of this hierarchy has a ‘node’ and nodes are aligned to correspond to separate timing slots. This means that place assimilations can be expressed by the deletion of an association to a place node and a reassociation to the place node of the upcoming timing slot. This reassociation process is generally taken to be motivated by a phonological rule. The example in Figure 1.3 shows an alveolar to velar assimilation.

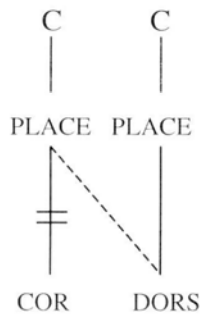


Figure 1.3 autosegmental representation of alveolar to velar assimilation

This model is ideal for anlaysing morphophonemic alternations such as in+come = income whereby the nasal is without doubt categorically velar. The prediction from the formalism in Figure 1.3 is of a discrete substitute of one feature value (DORS) for another (COR), resulting in the missing out of an alveolar articulation. While in this case, it is easy to see how this formalism might relate to a possible phonetic realisation it is important to disassociate it from the realm of actual physical gestures in the vocal tract. Modelling post-lexical assimilation as ‘surface phonology’ in this way is not really the standard domain of phonology which traditionally deals with lexical phenomena. The units that phonologists assign to tiers and nodes in autosegmental frameworks happen to have the same names as articulatory entities such as 'anterior', 'coronal' and 'voiced'. Phonological representations do not intend to model the vocal tract but do attempt to show the 'naturalness' of phenomena such as assimilation. Naturalness is a difficult concept to define but essentially arises from the 'economy' of phonological representations. It is generally accepted that it is easier and more desirable to fill in or add features than change them. Hence, coronals are underspecified in phonology because coronal is the default place of articulation for stops. Place assimilation is considered a 'natural' phenomenon because it is possible to arrange place features together on a hierarchy allowing place node association lines to spread, rather than +/- values changing.

Apart from the pitfalls of aligning phonological representation with phonetic facts, the overriding issue concerns whether, in fact, phonological models predict the correct phonetic outcomes. The instrumental studies discussed in section 1.3 below directly address this question.

1.2.1.1 Phonological underspecification

Underspecification theories have developed from Chomsky and Halle's (1968) proposal of markedness and redundancy among phonological features. Underspecification states that underlying representations of segments should be minimal and only consist of information about the distinctive properties of the segment. In 'radical underspecification' accounts this underlying information should include only unpredictable values for features whereas predictable, or default values come 'on-line' by rules during derivation (Archangeli, 1988). Keating (1990) suggests that underspecification is categorical at the phonological level but continuous on the surface level. In terms of place of articulation, underspecification theories view *coronal* as the least specified universally and so no place specification is required for coronals. Paradis and Prunet (1991: 182) explain that:

> The use of underspecification with a default feature-filling rule amounts to extracting the most frequent value of a feature for a given class of segments and building a bias into the language system to use that value of the feature unless it is specifically contradicted by other phonological information.

The fact that in many languages alveolars assimilate to other places of articulation, while velars and labials do not assimilate is taken as a major piece of evidence that coronals are underspecified for place of articulation in underlying forms. However, while underspecification simplifies lexical entries, at the same time it increases the amount of on-line processes which must act to fill the missing feature and thus overall brings no economy to the phonological system. In the case of alveolar place assimilation the exact nature of the on-line process needs further investigation through instrumental studies.

1.2.2 Articulatory Phonology

Browman and Goldstein (1990, 1992) have provided a radical account of speech processes whereby disparate phenomena, traditionally requiring separate phonological treatment, are unified under a single mechanism. Chiefly by a process of gestural overlap, sometimes in combination with a decrease in gestural magnitude, assimilation, deletions and weakenings occur. Browman and Goldstein argue that the primitive unit, called the 'gesture,' is the same for phonological and phonetic representations. However, Keating (1990: 48) has commented that 'even this theory needs phonetic implementation just as Targets and Interpolations models do , in the form of postlexical adjustments to the pre-specified gestures and alignments, and a decision mechanism for allocating gestures to articulators (called Task Dynamics).'

Linguistic (phonological) structures in Articulatory Phonology are represented as coordinated articulatory movements, gestures, which are themselves organised on a gestural score in the style of autosegmental representation (gestures are also argued to have perceptual relevance). Gestures are the aspect of the model most obviously aligned with what we think of as phonemic targets. They are relatively abstract and, importantly, invariant, being made up of coordinated goal-based movements. But even these movements are still abstract relative to the speech output signal. The output is generated from the dynamic model 'task dynamics' (Kelso et al, 1986) which provides three-dimensional articulatory geometry of a highly constrained nature. Fowler and Saltzman (1993: 172) provide a useful definitive statement on the difference between a gesture and a movement:

> ...[we] use the term gesture to refer to a "member of a family of functionally equivalent articulatory movement patterns that are *actively* controlled with reference to a given speech-relevant goal (e.g. a bilabial closure)" (p.334). [Saltzman and Munhall, 1989] According to this usage, gesture and movement have different meanings: although gestures are composed of articulatory movements, not all movements can be interpreted as gestures or gestural components. For example, when the

> vertical distance between the upper and lower lips changes due to the active coordination of the lips and jaw to produce a bilabial closure, the resultant movement pattern is considered to be a gesture. However, when the interlip distance changes as the passive consequence of the jaw's active participation in a different gesture (e.g. an alveolar gesture), the bilabial movement pattern would not be called a gesture

The fundamental principle behind this dynamic model which works in tandem with gestural phonology is that the complexity of speaking lies in controlling the coordinated movement of sets of articulators and not, as is often assumed, movement of individual articulators.

Vocal tract 'variables' are controlled by the task dynamics and they refer to the specified articulators required to accomplish speech goals. For example the gesture of lip closure is a single goal for the tract variable of lip aperture. Lip closing gestures involve the jaw, the lower lip and the upper lip which are precisely and functionally coordinated to achieve closure. Rather than each of the three articulators being assigned individual goals for labial closure, coordinative sets define the gestures that tract variables can achieve, lip closure being but one possible gesture from the variable lip aperture. There are several tract variables. These and their associated articulators are shown in Table 1.1 below, taken from Browman and Goldstein (1990: 344):

Table 1.1 tract variables and associated articulators (from Browman and Goldstein, 1990, 344)

tract variable		**articulators involved**
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict. location	tongue tip, body, jaw
TTCD	tongue tip constrict. degree	tongue tip, body, jaw
TBCL	tongue body constrict. location	tongue body, jaw
TBCD	tongue body constrict. degree	tongue body, jaw
VEL	velic aperture	velum, glottis
GLO	glottis	glottis

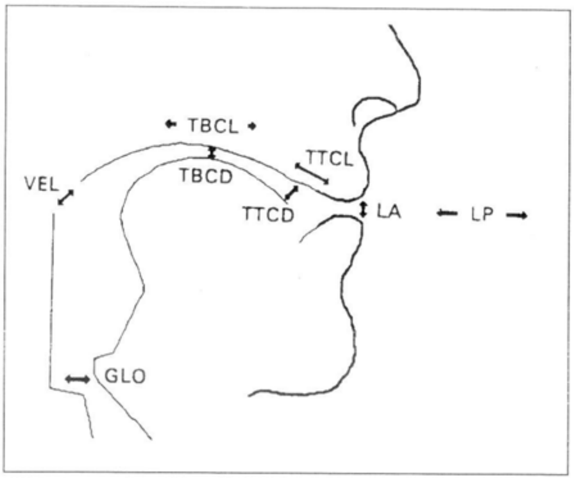


Figure 1.4 Tract variables and associated articulators (from Browman and Goldstein, 1990: 344)

The oral tract variables are grouped in 'horizontal-vertical' pairs: lip protrusion and lip aperture; tongue body constriction location and degree; tongue tip constriction location and degree. The non-oral variables are glottal state and velic aperture. Sets of movements combine to create a goal for tract variables, for example the upper and lower lip and jaw (which form the tract variable of lip aperture) implement the gesture lip closure. This is the way in which the complex task-dynamically modelled movements become linguistically significant. The structure which 'plots' gestures in relation to each other is known as the gestural score. So while coordinative sets of movements define a gesture, sets of gestures define an utterance. Gestures are organised into *tiers* on the gestural score which reflect the relative 'independence' of their articulatory function. Velic, and to a large extent, glottal gestures function most independently and occupy their own tiers. The articulators that are involved in velic and glottal gestures are not used to produce other gestures and so they are considered to form separate subsystems. Tongue tip and tongue body form separate tiers along with the lips; the jaw is an articulator shared by all (labels given to tiers and to individual articulators should not be confused). Browman and Goldstein consider their tier organisation to match with traditional organisational schemes, they are said to correspond to the conventional phonetic groupings of place of articulation - labial, lingual and dorsal, and also to autosegmental phonological tiers.

The oral tiers of lips, tongue body and tongue tip are each associated with a separate pair of tract variables, for example the lip tier accommodates lip protrusion and lip aperture while the tongue tip tier accommodates tongue tip constriction location and degree. Of course, the associated articulators for these tract variables are different. For instance as shown in

Figure 1.4 above, the tongue tip horizontal and vertical variables are driven by the articulators tongue tip tongue body and jaw. This means that the jaw and tongue body articulators are shared by the tongue body and tongue tip tiers, a situation which Browman and Goldstein interpret as making the two tiers a further subclass.

Gestures have two critical characteristics, abstractness and discreteness. Continuous non-linguistic interactive movement is underpinned by sets of discrete gestures which are simultaneously occurring. Gestures are abstract, discrete and invariant across different contexts but because they are spatio-temporally defined, they are free to overlap in time. Rather than coarticulation and allophone variation being seen as an automatic consequence of articulators of varying densities, speeds and inertias, especially in movement toward and away from targets their treatment in Articulatory Phonology posits them as an automatic consequence of overlapping invariant gestures which underlie all (or most) motor activity. There is no longer the problem of mapping static, discrete and invariant entities onto context-adjusted, dynamic speech signals, since gestural organisation can represent both categorical and gradient information.

Connected speech processes and particularly casual speech assimilation is handled mainly by gestural overlap. The question of how gestural *reduction* fits into this model is a little harder to discern. Figure 1.5 shows the different overlap/reduction outcomes. According to Browman and Goldstein (1990 :366), it is possible that gestures can be considerably reduced in magnitude even to the extent of deletion. But crucially, 'even deletion…can be seen as an extreme reduction, and thus as an endpoint in a continuum of gestural reduction, leaving the underlying representation unchanged.'

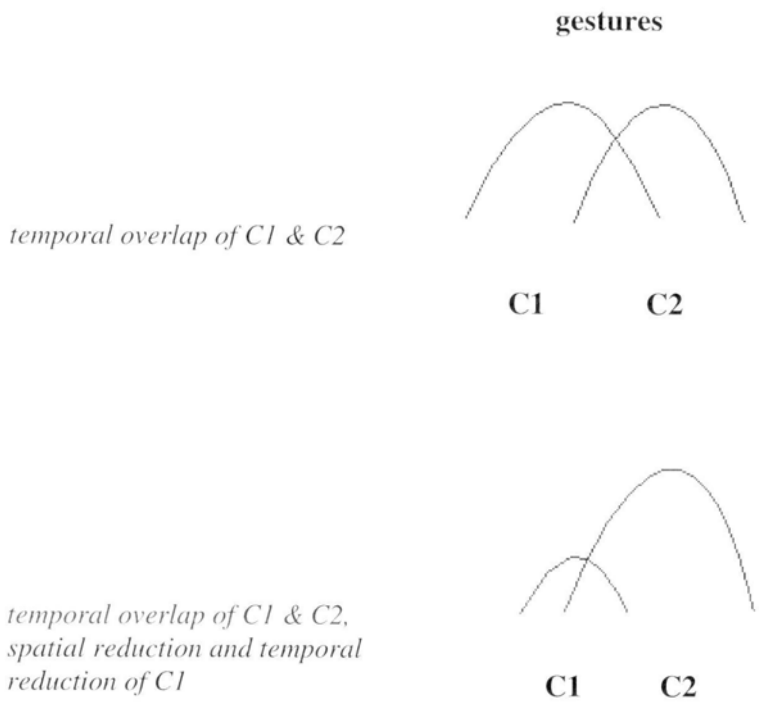


Figure 1.5 different combinations of overlap and reduction for two adjacent consonants

With respect to overlap outputs, Browman and Goldstein claim that gestures arranged on *different tiers* can be executed without disrupting the trajectory of each other i.e. they can be relatively independent, while still overlapping in time. This is permissible since they affect independent vocal tract variables and gestures are discrete. Ladefoged (1990) comments that the assigning of articulators to specific tiers on the gestural score is problematic:

> It has long been known that, for example, movements of the tip of the tongue affect the body of the tongue as a whole. Stevens et al (1986) have shown that such interdependencies can have interesting phonological consequences...it would now seem preferable to consider movements of the tongue, whether of the tip or the body or the root, as being deviations that could affect the entire shape.

Place assimilations in casual speech which involve gestures on separate oral tiers are, apparently, the most common types. It is believed that the most common ones, for RP at least, involve coronals assimilating to labials or velars (Brown, 1990 and Gimson, 1989). Browman and Goldstein take from Brown's separate lists of cases of assimilation and deletion, an example of each to illustrate that, as they see it, the cases of deletion are only apparent and that both deletion and assimilation are caused by overlap. For the case of consonant deletion they take the phrase *must be* /ˈmʌst bi/ → [ˈmʌsbi]. They predict that the alveolar constriction for /t/ is still present in the 'fluent' version of *must be* only it has been 'hidden' or completely overlapped by the bilabial closing gesture (Figure 1.6 is a schematic representation adapted from Browman and Goldstein, 1990: 361). This gestural blending effect occurs through the sliding in time of gestures on the gestural score. The fact that there would be no acoustic correlate of the /t/ is due to both alveolar closure and release occurring during the bilabial closing phase:

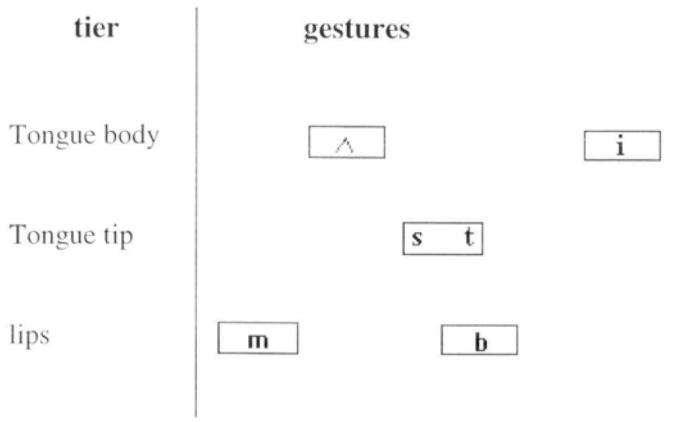


Figure 1.6 Hypothetical gestural score for oral tiers of 'fluent form' of must be (adapted from Browman and Goldstein, 1990: 361

Another example of theirs is how the perceived deletion of the alveolar in the phrase *perfect memory* is not borne out by the articulatory facts - the /t/ gesture is fully present but overlapped by the /m/. The bilabial closure is simultaneous with the /t/. Figure 1.7 shows their x-ray microbeam data for this utterance. The top panel shows *perfect memory* spoken in a word list and the lower panel shows these words spoken in a phrase. With regard to hidden gestures, they show that in a case such as *perfect memory* the citation form of the phrase and the connected speech phrase both include a tongue tip/blade raising gesture 'of roughly the same magnitude' while for the connected speech version the bilabial gesture /m/ overlaps the release of the alveolar gesture. For both versions, the velar closure is not released until the alveolar closure is formed - entirely discrete sequencing of /k/ and /t/ is probably quite rare. Importantly, Browman and Goldstein state that hiding is more likely to take place if the following word begins with 'true' consonants: 'in general…the more extreme the constriction associated with a gesture, the better able that gesture is to acoustically mask another gesture with which it is co-occurring' (1990: 366)

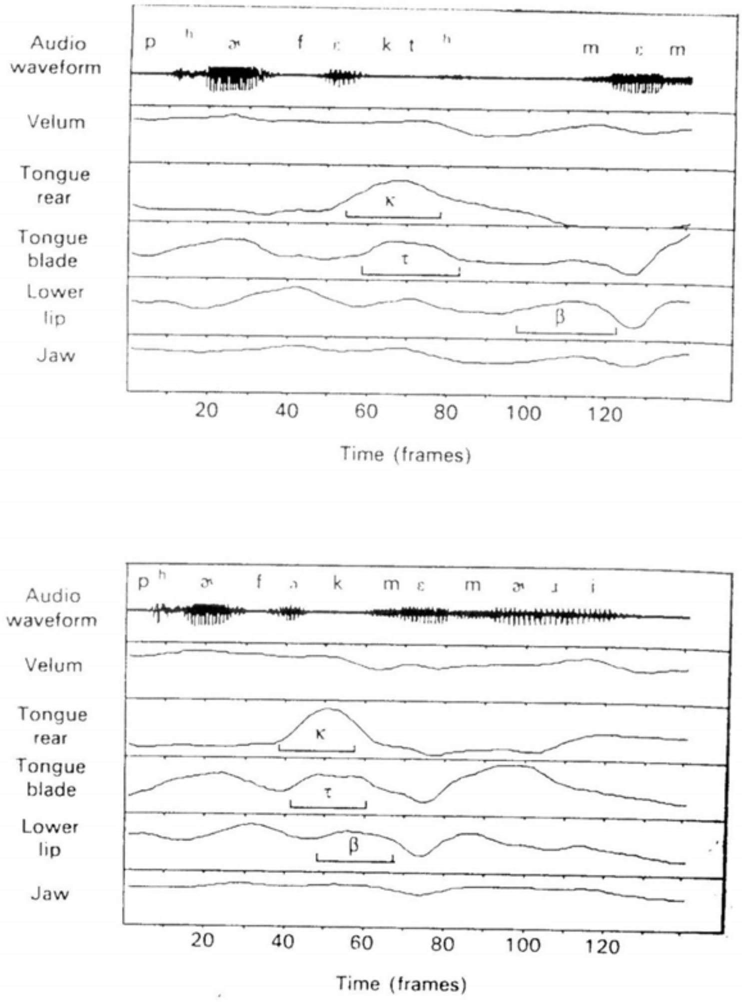


Figure 1.7 x-ray microbeam trajectories for 'perfect memory' *spoken in a word list (top panel) and spoken in a phrase (lower panel) taken from Browman and Goldstein, 1990: 364)*

Stetson (1951) had already noted these hidden gestures in the production of the nonsense syllable /ispda/. In careful speech all the consonants were kept distinct but in fast speech he noticed that the two alveolar gestures coalesced and overlapped with the bilabial closing gesture. The result was an articulatorily observed closure phase on a kymogram but no acoustic record of it.

A similar case of hiding is proposed for consonant assimilation. The example this time is *hundred pounds* /hʌndrəb ˈpaʊnz/. The same prediction is made as for the *perfect memory* example, the bilabial closing gesture overlaps the alveolar gesture while taking on the voicing properties of the /d/ resulting in the transcription [bp]. Again, as for *must be*, the alveolar gesture is presumed to remain intact although inaudible due to the phasing of voicing gestures. Traditionally, consonant deletion and consonant assimilation have been subdivided under general processes for connected speech but Browman and Goldstein hypothesise that they do in fact share the same source - increased gestural overlap between gestures on separate oral tiers. However, they concede that gestural *reduction* also has a part to play in connected speech processes of this kind.

In the case of consonantal weakening or lenition e.g. *must be* realised as [mʌsβi], the explanation given is of a decrease in the magnitude of an individual gesture. 'The reduction in amplitude of movement associated with the gesture then leads to the possibility of an incomplete acoustic closure'. As a rare comment on the *relationship* between overlap and reduction they say (1990: 370):

> Additionally, reductions in magnitude may combine with increased overlap, leading to greater likelihood of a gesture being "hidden" by another gesture...Such reductions of movement amplitude *often, but not always* occur in fast speech (Lindblom 1983; Munhall, Ostry, and Parush, 1985; Kuehn and Moll, 1976; Gay 1981). It may also be the case that gestural magnitudes are reduced simply because the speaker is paying less attention to the utterance (Dressler and Wodak 1982; Barry 1984). [italics added]

While the concept of overlap is clearly motivated by the contraction in time, corresponding to the horizontal axis of the gestural score, the mechanism behind gestural reduction is less clear. A phonetic explanation for reduction is, however, offered by Saltzman and Munhall (1989). They state that in the task dynamic component, the vowels that surround, for instance, an alveolar to velar sequence, share the same tract variable representation (tongue body) as the velar gesture. Thus the tongue body parameter is dominant in this VC_1C_2V sequence at the expense of the coronal gesture and this process, they claim, can still be considered gestural blending.

Finally, there are output predictions for adjacent gestures on the *same* tier on the gestural score. The prediction for these is not the same as for gestures occurring on different tiers. This is because it is unlikely that adjacent gestures, with different 'targets', can overlap without disrupting each others path e.g. *come from* /kʌm frəm/ → /kʌɱ frəm/. This is in contrast with gestures which do not occupy the same tier and which do not compete for articulatory time and space. The result is the type of blending which 'shows itself in spatial changes in one or both of the overlapping gestures' (1990: 362).

Through these examples Browman and Goldstein clearly make the point that speakers have limited stored (phonological) information which can change categories by rules. In this framework the business of mapping the abstract and the physical in speech (the most typical model of which from Levelt is described in section 1.4) is circumvented since these two entities are made of the same material, gestures.

Browman and Goldstein also comment on the relationship between assimilation and language-specific syllable constraints. In VC(C)(C)V sequences they propose that the consonant gestures are phased in relation to the immediately preceding one, so long as this does not violate the strict syllable constraint of the language. In this way, only those consonant clusters which could never form syllable onsets or codas such as *perfect memory* or *hundred pounds* can lead to complete gestural overlap/hiding. Underlying alveolar - bilabial or alveolar-velar gestures could never appear at syllable onset whereas labial-alveolar or velar-alveolar sequences could form syllable codas (e.g. *top layer / plank* and *pick lemons / clock)*

Kohler (1992) considers how Articulatory Phonology can account for all cases of gestural reorganisation. In particular he observes that often the very common final syllable [ən] is reorganised so that the preceding oral closure is not released but is kept continuously until the offset of the final nasal. This usually results in [m] or [ŋ] after labials and dorsals without a trace of an apical gesture. Articulatory Phonology dictates that consonant place assimilation is restricted to immediately adjacent consonants as a result of gestural overlap, however, Kohler points out that in German the restructuring of gestures can involve more than two units as in *mit bunten Papierschlangen* [...'bʊmpm p...] (with coloured streamers). This would be difficult to explain in terms of simple gestural overlap 'here the speaker has to plan a gestural reorganisation a long time in advance of the focal gesture of [p] at the beginning of the last word, i.e. before the onset of the first nasal articulation.' (p.207) This contradicts the approach of Articulatory Phonology that in fluent speech no lexical phonological gestures are ever transformed into other gestures. He gives another example in German: *Beamten* [bə 'ʔampm] (civil servants) where the labial closure is sustained across a nasal-stop-nasal sequence and the stop formation and release are controlled by velic movement alone. Kohler says that in this case of total assimilation it is inadequate to say that the 'chaining of the lip, tongue tip and velic aperture gestures produces certain changes in these tract variables.' Most interestingly though, he comments that Articulatory Phonology 'lacks a functional organisation of gestures in phonology in place of a purely anatomical one. And this deficit is at the root of why - in spite of its great explanatory power - it falls short, in many cases, of giving principled reasons as to why reductions reach the points that are observed and do not go further or stop short elsewhere.' (p.206)

Ohala's view (1990) on the domain of phonetics and phonology contrasts strongly with the Articulatory Phonology framework. He sees phonetics and phonology as separate for functional reasons. He regards autosegmental notations as effective for describing and classifying sound patterns which physical phonetic representations cannot. In particular, the type of phenomena autosegmental notation was invented to handle is best served by it: certain phonological 'melodies' which are abstracted out of prosodic properties of words like tone and quantity. Traditional phonological accounts are also good for things like tracing the phonological events leading up to current sound patterns. Phonetics, however, is good for pin-pointing the initiation of sound change. Ohala has no problem with mixing modes of description if appropriate and this is typical of his resourceful approach to speech. He draws on laws of physics and nature to show that some ad hoc 'patches' onto existing phonological frameworks, in fact, will not always be peculiar to speech: 'If Boyle-Mariotte's Law has to be invoked to explain the devoicing of stops, it is the same Boyle-Mariotte's Law that applied to other parts of the familiar universe we live in (bicycle pumps, automobile pistons, party balloons, barometers, etc.) (Ohala, 1990: 267) This he says, is in contrast to currently popular phonological representations which enhance sound patterns by 'conjuring up a vast array of devices and conventions that seem to apply exclusively to speech.' He proposes that there are certain known physical principles (phonetic primitives) which can be effectively bound to existing representations and that a unified formulation which has the power to explain why assimilation is uni-directional is a distant prospect. Specifically, the primitives that can be incorporated are aerodynamic principles, principles relating to vocal tract shape and acoustic output, and principles of how the auditory system derives information from the acoustic signal. The classificatory use of square-bracketed entities such as "obstruent", "labial" and "oral" and the binary oppositions they are able to form cannot show 'how this combination will create an increase in oral pressure which, when released, gives rise to a rich set of place cues.' (p.268) See section 1.5 for a discussion of acoustic consequences of consonant position when in assimilatory contexts.

In general, Articulatory Phonology presents a useful framework for expressing various low-level phonetic effects that can take place when gestures come into contact with each other.

The gestural score represents high-level phonological information while the task-dynamic model formulates actual gestural trajectories which have the potential to generate low-level effects. In the case of alveolar to velar sequences, the principle phonetic effect is gestural overlap which ultimately may 'mask' the coronal gesture so that a complete assimilation is heard. However, where Articulatory Phonology appears compromised is in its account of gestural weakening. It is difficult to see how gestural weakening arises from the properties of the task-dynamic model without a re-programming of the execution of the coronal gesture. More seriously, there is no opportunity in the model to express coronals as phonologically special items because all gestures are treated as formally identical entities.

1.2.3 The Window model of coarticulation

Keating's Window model of coarticulation (1988: 1990) is an attempt to account for the continuous changes in time and space during speech. The model challenges the traditional phonetic view that these graded effects are merely due to automatic and inevitable consequences of the physical properties of the speech production mechanism itself. Instead, Keating's model attributes coarticulatory effects – those thought to be automatic and universal and those thought to be phonological in origin – to the level of phonetic rules within the grammar.

In this model it is the interpolation between 'windows' which mirrors the smooth spatial trajectory the articulators describe between the physical values of adjacent segments. For a feature value associated with a given articulatory or acoustic sequence, a range of values called a *window* is assigned. On the whole, though, windows depict physical dimensions and their interpolations.[1] Adjacent windows are linked by 'paths' which interpolate between windows. The paths are governed by the principle of minimal articulatory effort. The width of the windows is the most crucial aspect here, which represents the entire range of possible contextual variability of a segment. Thus a feature value does not have a different window for different phonetic contexts, it has only a single window. Specified features have a narrow window and so contextual variation is restricted, but a maximally wide window reflects an unspecified feature and a large amount of contextual variation is

[1] Keating does acknowledge however that similar windows could be constructed for *acoustic* measurements such as formant frequencies since these are often utilised as physical representations although there might be a problem with the non-linearity of articulatory-acoustic mappings

possible in the output (see Figure 1.8 for some examples of windows and paths which will be discussed below).

Phonological information is required at the outset for this model. Keating stresses that 'the phonological feature values that are the basis for window selection need not be the same as the underlying values: phonological rules to change or spread feature values still apply before the phonetics. Thus in terms of segments, windows are selected for extrinsic allophones rather than phonemes.' (1990: 456). Like Articulatory Phonology, Windows places the 'continuous' aspect of speech production at the centre of the framework.

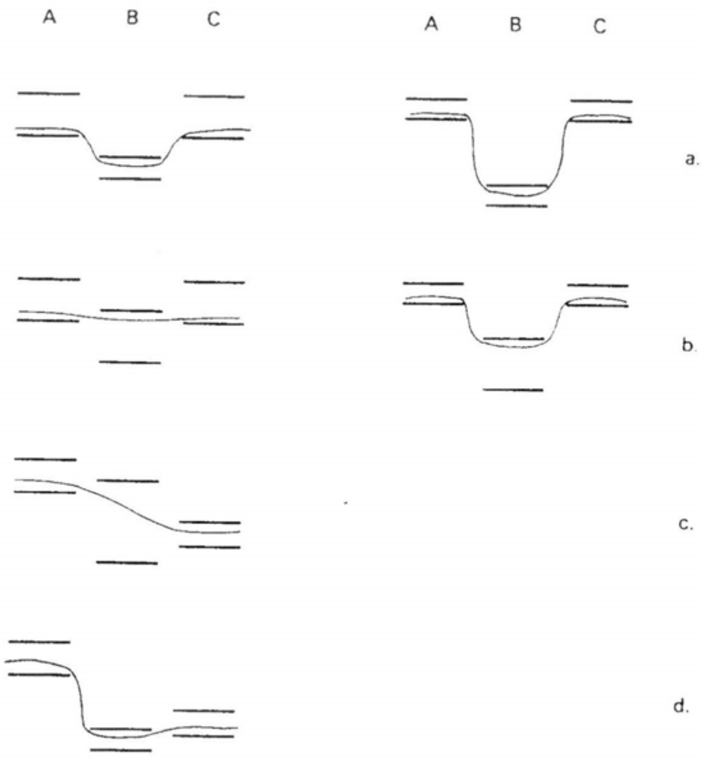


Figure 1.8 Hypothetical 'paths' derived from sequences of windows of varying widths (taken from Keating 1990: 457)

The examples in Figure 1.8 which Keating gives are designed to show the progress of a path along a single articulatory dimension for varying articulatory sequences. Only a subset of the sum of the possible physical range for the articulator is represented by the windows for each example. In Figure 1.8(a) she shows the effect of a medial narrow window between two identical surrounding windows. This middle window exerts a strong constraint on the interpolation line overall and affects the abruptness of the offset/onset interpolation within the adjacent segments. The middle narrow window itself, of course, shows that the feature it represents will vary little across contexts. In articulatory terms, the more specified and precise the formation of an articulatory target is (and hence the narrower the window) the more likely it is to require a sudden and 'steep' re-configuration of the articulators. But there is no explicit mechanism in Keating's model to associate this sudden discontinuity with timing or velocity as might be expected: 'I also leave aside the question of how timing fits into this scheme e.g. the time interval over which paths are constructed, and whether windows have variable duration, or are purely notional' (ibid.: 457). Another factor left unexplained is how windows represent phonetic variation due to different speaking styles. In the current model, informal speech exhibiting strong coarticulation /target undershoot and clear speech are not distinguished because window width represents all possible variation and conflates segmental and non-segmental effects.

In Figure 1.8(b), Keating considers sequences with the same surrounding windows as in (a) but this time with wide middle windows. The wide window here requires a less abrupt transitional turning-point and, as in left-hand figure, sometimes requires very nearly no turning-point. In Figure 1.8(c) a wide window comes between narrower windows. This middle window contributes nothing to the resulting path, allowing instead an unfettered and smooth route from one specified value to another. In 1.8(d) the narrow window between the same two unlike narrow windows as in (c) does makes a contribution to the path. On the subject of 'turning-points', Keating comments that they are not intrinsic to a segment. Determining turning-points in paths 'is divided between a mechanism which provides the windows, and the interpolation mechanism. The individual values associated with segments do not exist before an actual curve is built; there are no "targets" or assigned values.' (ibid.: 456)

A maximally wide window which manifests what looks like a complete lack of inherent phonetic value is at one extreme of a continuum, according to Keating, while a minimally narrow window is at the other where contextual variation is expected to be zero. There is a category called "not quite unspecified" which is represented in Figure 1.8(b) left-hand side. Here, the middle window is not so wide that it allows direct interpolation between the bordering windows. This middle segment is not completely unspecified. Keating considers English vowels with regard to nasality to fall within this category. Phonologically, these vowels are unspecified for nasality but she shows that they are associated not with maximally wide windows but just wide windows and in some environments, even display their inherent specification. This, she argues, occurs because underspecification is categorical at the phonological level but is continuous at the surface representation level. Keating admits that her examples are somewhat oversimplified. Furthermore, the relation between a feature and a physical dimension can sometimes be rather indirect and she gives the example of tongue and jaw position relating to more than one phonological feature, 'thus, phonetic implementation involves interpreting features as physical dimensions in a potentially complex way.' (ibid.: 456).

The most important property of the model, in terms of relevance to the issue of connected speech processes, is that phonologically underspecified features may persist into the surface representation. While traditional feature-spreading models always fill-in underspecified features with features from adjacent phonemes in a categorical fashion (thus ensuring the presence of a target), Windows leaves open the possibility that, in the case of coronals for instance, there may not be an inherent contribution to the physical path the articulators eventually take with regard to place of articulation. This leaves room for the gradual, 'more-or-less' type of variability in the amount of alveolarity that is produced, depending on the nature of the surrounding windows (phonetic context).

1.2.4 Gesture economy

In terms of 'low-cost' versus 'high-cost' speech production strategy, coarticulation has a particular relevance. If the speech motor system is regulated by the principle of economy, as other types of motor behaviour are said to be, coarticulation is the central mechanism. Lindblom's seminal study of vowel reduction (1963) showed that reduction should be viewed as an inevitable response of the speech motor system to an increase in the rate at which motor commands are issued. Rather than reduction being a phonological 'centralisation' process towards the vowel space around schwa, Lindblom believed reduction to be a coarticulatory process aimed at preserving energy. CVC syllable data was collected and it was found not only that acoustic targets were not always achieved but that the formant frequencies of the vowel at their mid-point changed as a function of vowel duration. As duration decreases, formant transitions in general shift toward the acoustic values of the consonant, a phenomenon Lindblom called target undershoot. This was interpreted as evidence that articulators do not have enough time to execute full displacements whenever multiple motor commands are issued in quick succession. Nord (1986), however, showed that this type of reduction can also occur at slow rates, meaning that speakers can to some extent select different production strategies according to the communicative situation.

The idea that this low-cost strategy in fact underpins normal speech, has been proposed by Lindblom, Pauli and Sundberg (1975). They found that in the production of VCV sequences, the best model of the tongue configuration for the consonant was one characterised by minimal displacement from the values of the surrounding vowels. Speakers favour a tongue body shape for the apical articulation involving an adjustment to the tongue body shape required for surrounding vowels.

A major problem with this economy of effort model is that it presupposes that all speech commands are issued serially. As a result it cannot account for the kind of anticipatory coarticulation that spans many segments. Sussman and Westbury (1981) for instance found that the onset of anticipatory lip rounding occurs earlier in antagonistic segments than in neutral ones for some sequences. This indicates that the speech production system builds in a mechanism to look ahead and plan articulatory simplifications which may mean a reversal of specified features.

Lindblom's later 'H-and-H theory' (1983, 1990) proposed that speech production is the result of a play-off between the speaker's own claim on speech economy and the listener-oriented demands on successful communication. The essence of this is that speakers can adapt their production to the perceptual demands of the listener and so when the communicative context requires a high-level of phonetic precision, over-articulation, *hyper* speech, will come into play. Conversely, when the context demands less 'clarity' speakers are able to exploit the option of economical articulatory effort and under-articulate, *hypo* speech. All speech production is supposed to fall somewhere on a continuum between hyper and hypo speech and at the hypo end of the continuum, listeners are said to pay more attention to signal-independent information in order to recover the message. Coarticulation has a very important role in this framework. In articulatory terms coarticulation gradually increases as production moves from hyper-speech towards hypo-speech and the perceptual concomitant of this is that a move towards hypo-speech is characterised by a gradual decrease in phonetic contrast. In this way the articulatory/acoustic properties of a speech item will exhibit a broad range of variation which reflects the hypo-hyper production continuum. In part the development of the H-and-H theory was a response to coproduction theories which concentrate exclusively on the kinematic properties of speech production while neglecting the role of perceptual output constraints.

On the subject of 'ease of articulation', to some extent linked to the principle of economy of speech motor activity discussed above, Ohala (1990) comments that 'ease' or 'simplicity' has never been adequately and convincingly defined. When a heterorganic cluster becomes a geminate (homorganic), or when a nasal stop assimilates in place to a following consonant 'there is one less articulator involved but it does not follow so straightforwardly that this yields an easier task'. But, he argues, 'no one knows how to quantify articulatory effort but certainly the neurological control operations should be counted too, not just the energy required to move the speech organs.' (p.260). To illustrate this point he gives an example found in Shona of nasal assimilation to the place of a following stop:

N + bato > mɓato

Here, the velum is having to close in the middle of the bilabial closure, which spans two segments, rather than combine this with the offset of one segment with the onset of another which has a different specification for place which, Ohala suggests, might 'cost more'. Similarly the extra duration of a geminate may cost more than a consonantal closure of normal duration. But what causes Ohala to be most sceptical of the ease of articulation account is that it fails to explain why in VCCV sequences it is C_1 and not C_2 which can assimilate. He says that if ease of articulation really was the origin of this phenomenon then once C_1 had been formed then lazy speakers would maintain it at the expense of C_2.

1.3 ARTICULATORY STUDIES OF PLACE ASSIMILATION

During the past thirty years, the technique of electropalatography (hereafter EPG) has been an important tool in the investigation of lingual aspects of place assimilation. Perceptual or acoustic approaches to this alone are not always revealing. Particular interest has been shown in the 'instability' of word final alveolar consonants when adjacent to word-initial velar consonants (in German, Kohler, 1976 and in English, O'Connor, 1973), a phenomenon that easily lends itself to analysis using tongue-palate contact tracking. Post-lexical alveolar to velar assimilations e.g. /stɑkˀ kauntɪŋ/ *start counting,* are said to be more frequent than alveolar to bilabial assimilations e.g. /tɛm mɪnɪts/ *ten minutes* (Gimson, 1989:310).

The major consensus from many experimental observations of the alveolar to velar assimilatory process, most of which will be discussed below, is that it is gradual and does not, after all, conform to the type of phonological formalism described in section 1.2.1. The term gradual refers to the fact that connected speech processes like assimilation are not applied by speakers in an all-or-nothing fashion. Assimilation is not either absent or complete, with complete assimilation motivated by a phoneme substitution rule. Rather, assimilation can be partial yielding intermediate forms between non-assimilations and complete assimilations. The observation of these intermediate forms has prompted the view that high-order phonological rules have either no part to play or have only a limited part to play in assimilation. Thus phonological theory and instrumental research have tended to locate assimilation within the abstract planning stage and the concrete physical execution stage respectively. There is some debate, however, within the instrumental research paradigm about the extent to which the properties of the speech production mechanism itself can account for observed assimilatory forms. These issues will be discussed in later sections. Other issues discussed include the effect of various phonetic factors on assimilation. Speech rate is probably the most significant of these (section 1.3.2 expands on this). Syntax is also a factor (Holst and Nolan, 1995) as is 'habituation'. Habituation refers to the phenomenon whereby an increase in the application of a connected speech process for a particular word is thought to be brought about by a speaker's familiarity with it (Dressler and Wodak, 1982). Assimilatory variability must also be seen in the light of variation both between and within individual speakers.

Keating (1996), amongst others have referred to evidence of gradual assimilation as 'phonetic assimilation'. Reminiscent of Jones' distinction between 'assimilation' and 'similitude'(1932), she says that assimilation can be phonological (categorical) or it can be phonetic (gradient). The latter is due to overlap, reduction or interpolation (articulators moving toward and away from targets) and is hence considerably more common than is generally realised. However:

> this does not mean that sub-phonemic assimilation can never be categorical. Feature-spreading (or similar phonological operations of assimilation) in principle can produce categorical allophones, that is, derived feature combinations with targets different from those of the underlying segments. Nonetheless, this seems to be a rare occurrence. When suspected cases are examined carefully, they generally show gradient characteristics. (p.55)

1.3.1 Partial assimilation

Early EPG studies confirmed that speakers produce /VtkV/ and /Vt#kV/ sequences with a variety of different spatial patterns including the all-important partial assimilations (Hardcastle and Roach, 1977). In this work, separate articulatory strategies were identified for the production of alveolar to velar sequences. These are: (i) clearly defined alveolar closure followed by a short period of simultaneous alveolar/velar closure; (ii) some indication of an alveolar gesture (although very short compared to a 'canonical' alveolar) with an incomplete occlusion at the alveolar ridge; (iii) velar stop pattern substantially further forward than the same subject's production of an /iki/ target sequence, with no contact in the alveolar region (what might currently be referred to as a residual alveolar) and (iv) a 'normal' velar pattern identical to an underlying velar. Figure 1.9 shows the tongue-contact EPG patterns for an example of one of their /t#k/ tokens, in the environment of a high front vowel. It would be classed as a partially assimilated token, type (iii). This palatal stop configuration can be thought of as an extreme example of a co-produced or 'blended' front-back sequence predicted by Recasens et al (1993) and discussed in more detail in section 1.3.4. Each numbered EPG frame indicates the location of tongue-palate contact (in this example the large unfilled squares indicate contacted EPG electrodes and the small dots represent uncontacted electrodes). Frames are 10ms apart. The bottom of the individual EPG frames represents the alveolar region of the palate and the top of the frames represent the velar region. The boxed frame 195 is where velar closure occurs in this example and the boxed frame 213 is where the velar closure is released.

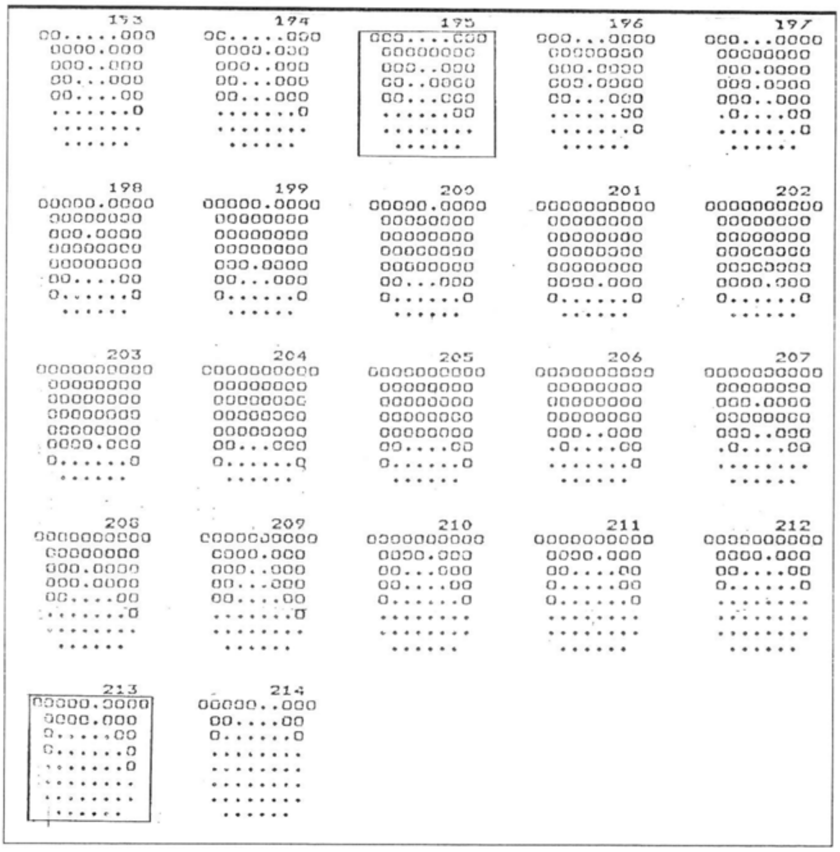


Figure 1.9 EPG printout showing a partial assimilation (palatal stop closure) for /t#k/ in a high-front vowel context: 'eat cake'.

Another study by Hardcastle (1994) looked at the production of adjacent alveolar to velar stops in natural connected speech. This work was concerned primarily with the phonetic factors that might constrain assimilation. Specifically, the focus was on whether the *type* of alveolar to velar sequence is a determining factor in their occurrence, that is, whether the alveolar is a nasal or a non-nasal or is part of a cluster. The distribution of residual articulations amongst and between individual speakers was not of direct concern. The test items captured the following potential assimilation sites:

/d#k/ → /g#k/ *Fred can*

/n#g/ → /ŋ#g/ *can go*

/n#k/ → /ŋ#k/ *Susan can't*

/nt#g/ → /ŋk#g/ *can't go*

Seven subjects repeated each of these five times. The data fell into four different EPG contact types: velar (complete assimilation); palatal ('partial' alveolar); alveolar plus simultaneous velar/palatal (double articulation); alveolar followed by velar/palatal (sequentially discrete). Interestingly, the pooled results showed that the experimental sequences generally motivate different assimilatory tendencies. Overall, however, there were far more assimilated (partial and complete) alveolars than non-assimilated alveolars although this was less in evidence for the /d#k/ sequence which suggested that nasals may be more susceptible to assimilation than non-nasals. There was a clear preference for subjects to assimilate the nasals in both the /n#k/ and the /n#g/ sequences but 'the /nt#g/ *can't go* sequence was complicated by the presence of an underlying [t] manifested by most speakers as a glottal stop, although one speaker had a partial alveolar contact accompanying the glottalisation.' (p.66). Attention is drawn to the fact, however, that non-nasal assimilatory contexts appear in comparatively stressed position in the experimental sentences while /n#k/ and /n#g/ occur in comparatively unstressed positions. It also emerged that when the abutting velar is *voiceless*, the incidence of unassimilated alveolars increases. There were some interesting results concerning intra-subject variability, assessed by observing firstly, differences in strategy between the speakers' 5 repetitions for a single sequence and secondly, by observing differences in strategy for the different sequences. For the former assessment a fairly high level of consistency was identified for most subjects. Consistency in production for the 5 /d#k/ repetitions was a different story, however, where 4 of the 7 subjects produced a mixture of assimilations and non-assimilations. Unfortunately, no information is given about how many of the assimilations were residual. More subtle intra-subject variability, of the order [əŋ] instead of [ŋ] for *can* was reported however. All subjects except one produced at least one assimilated sequence across all sequences.

Some studies have suggested that speakers do have some freedom in coarticulatory behaviour beyond that which can be attributed to anatomical variation between speakers' vocal apparatus. Nolan (1985) found inter-speaker variation in the coarticulation of English /l/ and /r/ with following vowels and Lubker and Gay (1982) noted differences between speakers, some Swedish and some English, with respect to lip rounding. Some speakers anticipated lip movement according to how much time they had available to do this, and thus could be said to implement a look-ahead type of coarticulation mechanism, whereas other speakers commenced lip movement at a constant time regardless of maximum time available. They also noted language-specific effects in this experiment whereby type of strategy tended to follow a particular language. More discussion of language-specific aspects of coarticulation/assimilation can be found in section 1.4.

Hardcastle and Barry's study (1989) lends further support to the view that post-lexical variations are gradient in nature. They explored the process of /l/ vocalisation of six speakers with differing regional accents using EPG, a process thought to be applied somewhat inconsistently amongst speakers (Wells, 1982). In common with Hardcastle (1994), the primary purpose of the experiment was to establish some articulatory/phonetic factors which might influence the application of it. In particular they address the hypothesis that regularities in the vocalisation process would be attributed to physiological and perceptual constraints overriding any dialectal effects. For this reason this study is also less concerned with the distribution of partial assimilations within and between speakers.

Materials for the experiment were designed to affect the components which define 'dark' /l/, that is, central alveolar contact of the tip and/or blade of the tongue and the raising of the back of the tongue towards the soft palate. In EPG terms /l/ typically shows a full occlusion across rows 1-3 and varying amounts of lateral contact depending on the vowel environment. The auditory correlates of these are a 'lateral liquid' impression and a half-close to half-open back vowel quality. The materials consisted of post-vocalic clusters containing /l/ in stressed sentence-final position. The /l/s were followed by consonants with differing degrees of frontness and tongue configuration and preceded by front and back vowels. Some examples of the stimuli are shown below:

1. When was the 'house 'built?
2. The 'car 'drew to a 'halt.
3. In 'some areas they speak 'Welsh.
4. 'Have you 'spoken to Mr. 'Walsh?
5. 'Could you 'fetch a 'pint of 'milk?
6. We 'buy the 'seeds in 'bulk.

Each subject read the sentences twice at normal conversational rate. On the basis of EPG records, /l/ tokens were judged + vocalised or -vocalised (i.e. made with alveolar contact or not). Problematic tokens were those where the alveolar /l/ configuration was not complete. Partial contact with as few as two or three electrodes contacted in the alveolar region was quite common. These were recorded as non-vocalisations because 'they are categorically different from a true vocalised /l/ in that a build-up of contact from the back to the front of the palate does not occur' (p.10) Out of a total of 102 non-vocalised /l/ tokens, 12 involved partial alveolar contact and it is interesting to note that 11 of these preceded sibilants. Hardcastle and Barry characterise this as the result of an anticipatory grooved tongue configuration during the apical /l/ gesture before /ʃ/. Further evidence of contextual phonetic influence on /l/ vocalisations is provided with the result that one speaker had consistently incomplete /l/ articulations in 'hills' and 'halls' but complete vocalisation before the postalveolar fricative in 'Welsh' and 'Walsh'. Another speaker had full /l/ alveolar contact for 'hills' and 'halls' but incomplete contact for /l/ before the postalveolar fricatives. Figure 1.10 shows an example of one of these tokens:

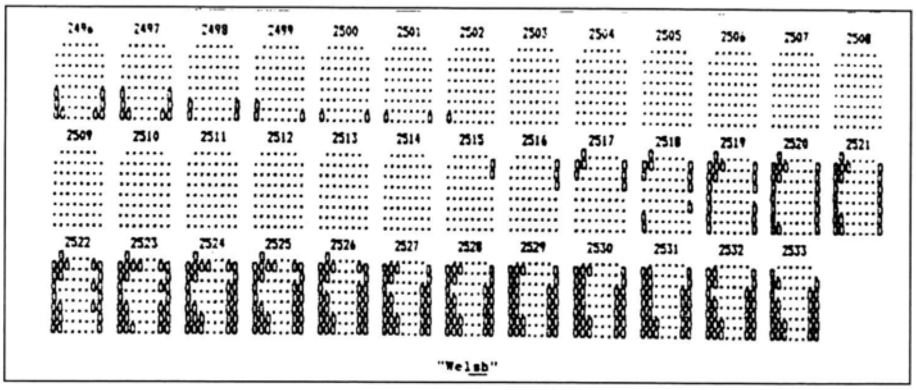


Figure 1.10 EPG patterns showing the /l/ cluster in 'Welsh' – *partial alveolar contact for /l/ (from Hardcastle and Barry, 1989)*

The authors comment:

> The phenomenon of partial alveolar contact has important theoretical implications in suggesting that /l/ vocalisation is not a categorical process. The apical gesture shows signs of extreme undershoot with some evidence of anticipatory coarticulation with /z/ and /ʃ/. It seems therefore that the vocalisation process should not be seen simply as the suppression of the apical gesture under certain conditions but, speaking figuratively, as a reduction in the intensity of the apical command. Under certain coarticulatory conditions e.g. before alveolars, this can result in full apical /l/ contact, while in the context of back consonants, total vocalisation may occur. (p.15)

There was a general tendency for vocalisation to be more frequent in the context of front vowels than with back vowels and this is explained by means of a perceptual framework. In vocalisations, the close or close-mid vowel (which constitutes the velar articulation of [lˠ]) is more of a contrast with front vowels than back vowels and so the contribution of alveolar contact for /l/ becomes less imperative. In this way, a vocalisation in 'milk' is more likely than in a word like 'bulk'. Although speakers only produced two repetitions of each experimental sentence, subjects were consistent in their realisations of /l/. In only 4 out of 72 cases was /l/ vocalised in one repetition and not in the other.

Speculation that a phonetic residue of an underlying alveolar gesture will always be present in the articulatory/acoustic signal, even in apparent complete assimilation, had gathered momentum as increasing evidence of 'partial' or 'residual' assimilatory forms has come to light. Dinnsen (1985:276) took this idea to its furthest conclusion:

> Every genuine phonological distinction has some phonetic reflex, though not necessarily in the segments which are at the seat of the distinction. Two utterances which are identical phonetically must also be identical phonologically.

The literature on phonetic neutralisation also explores this issue and can be paralleled with the ideas about assimilation covered throughout this dissertation. German /b, d, g/ and /p, t, k/ are said to be all realised as voiceless in syllable-final position. Many studies of different languages have investigated the possibility of surface traces of underlying contrasts which might indicate that neutralisation was incomplete. Chen (1970), amongst others, found in a cross-linguistic study that vowels preceding surface voiceless obstruents were 20% longer before underlying voiced obstruents than the voiceless equivalents. However, the method of eliciting data through subjects reading orthographic forms which make underlying forms apparent, and thus more likely to result in the preservation of contrasts, has been criticised. Another question mark hangs over the fact that evidence of incomplete neutralisation appears for only some subjects. Nolan (1986) has pinpointed the wider implications of neutralisation for phonology and the production/perception relationship. He points out the difficulty of reconciling neutralisation, attributed to a high-level feature changing rule, with the assumed low level phonetic implementation differences of the kind mentioned above. The ultimate solution to this problem of deciding where phonological rules meet phonetic implementation is, of course, to throw out any reference to phonology at all. Because neutralisation/assimilation is incomplete the phonological structure remains intact and thus the only task is to order the implementation tasks accordingly. But:

> This view then raises the more interesting question of whether the phonetic reductions caused by implementational short cuts are simply gradual and progressive up to the point at which phonetic identity is arrived at in the realisation of distinct underlying forms; or, instead, once a threshold of reduction is crossed, phonological restructuring intervenes to modify phonological forms (either at the level of underlying representations, or within the phonological derivation). p.7

This preliminary idea manifests itself in full experimental form in a study of assimilation by Holst and Nolan (1995), the details of which are given in section 1.3.5.

In terms of representation for transcription purposes phone-type segmentation implies temporal discreteness and paradigmatic discreteness, clearly not a helpful framework in this situation. On the subject of implications of incomplete processes for the production and perception relationship, Nolan cites (limited) evidence that these phonetic residues can be utilised by listeners to recover underlying structure. Port and O'Dell (1985) have shown that listeners can recover underlying German forms from incomplete neutralisation at a rate of 59%, although Nolan also cites Costa and Mattingly (1981) who showed that the residual vowel duration difference in New England *cod* versus *card* (a pair whose distinctiveness is in the process of being lost) is not picked up by listeners. The possibility that speakers are maintaining underlying forms in their phonetic output but listeners are hearing collapsed phonemic distinctiveness has great importance for Nolan: 'we are faced with the troubling prospect that each individual operates with distinct production and perception phonologies.' (p.8).

Wright and Kerswill (1989) take up the issue of the perceptual correlates of incomplete assimilation in an EPG /perceptual study of alveolar assimilation. Another concern is again whether or not the articulatory details of this type of assimilation support or contradict the 'textbook' phonological modelling of the process which posits it as always discrete. The aim of the perceptual part of their experiment was to ascertain whether trace alveolars remain in apparent complete alveolar assimilations (velar stop EPG patterns) that appear to be identical to lexical velars. Data was acquired from a professional phonetician who was asked to produce the experimental sequences with varying degrees of assimilation. The sentences were spoken at a constant rate and with a consistent intonation pattern. Sentences of minimal pairs were devised, of the type: *road collapsed/ rogue collapsed* and *did gardens/ dig gardens* where potential alveolar assimilatory forms can be compared with consonantal sequences which contain the underlying velar form.

They define four articulatory types from the data: full alveolar; (complete closure across one or more rows at the top of the EPG display) residual alveolar; (lacking mid-sagittal closure at the top of the EPG display but with side contact along the outer sides of the EPG display, also known as 'lateral' contact, a minimum of one row further forward than in the speaker's control velar-velar sequences), zero alveolar (complete assimilation indistinguishable from all-velar controls) and non alveolar (all-velar control sequences). Figure 1.11 shows their sample data.

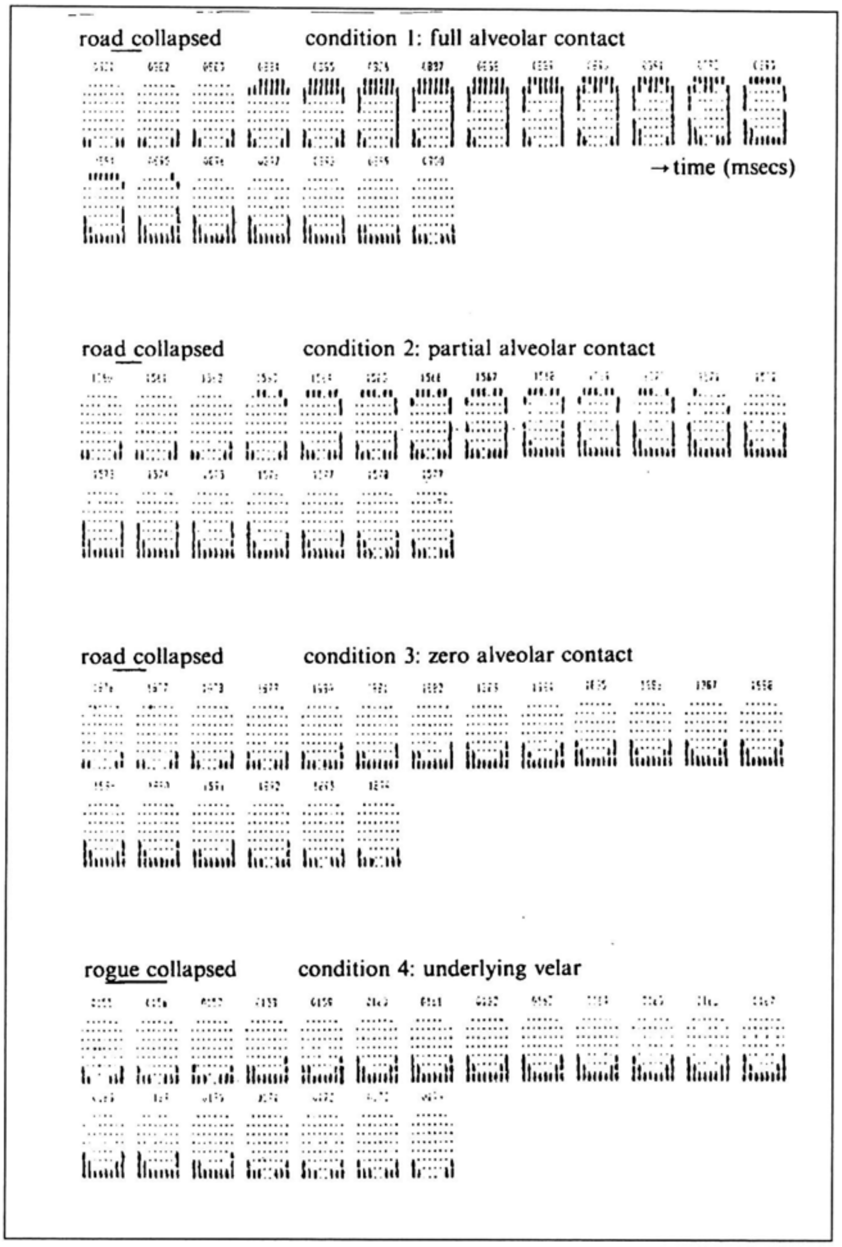


Figure 1.11 EPG articulation types for non-assimilated, partially assimilated and completely assimilated alveolars and for a lexical velar (taken from Wright and Kerswill, 1989: 51)

Wright and Kerswill do not refer to these EPG-defined intermediate forms as 'partial assimilations'. They take care to use the label 'alveolar' in their description of all articulatory types they found. They avoid the term 'assimilation' in order to reinforce their hypothesis that underlying alveolars are always manifested in some form articulatorily.

Fourteen trained phoneticians perceptually assessed the data acquired from Wright and Kerswill's single informant. What they found was that alveolars, realised as complete stop closures, were not surprisingly, highly reliably identified as such. In a series of paired *t-tests* the difference between the alveolar-judgement responses for the full alveolar versions and for the partial alveolars was statistically significant as was the difference between the responses for the partial alveolars and the zero alveolars (although there is potentially a big articulatory difference between these two types in view of the fact that patterns showing alveolar contact but with no complete closure are included in the 'partial alveolar' group). But this was not the case for zero alveolars and non-alveolars. For the perceptual judgment of the zero alveolars, only 40% were correctly identified as such, while 55% of responses were of partial alveolars. Similarly for the controls a very high proportion of responses were made up of partial alveolars and zero alveolars. However, the actual alveolar-judgment scores for each articulation type (full alveolar, residual alveolar, zero alveolar and non-alveolar) reveal, the authors propose, that there is in fact a perceptual continuum corresponding to the four articulation types; alveolar identification scores decrease almost linearly across the conditions.

Wright and Kerswill focus on this drop in alveolar judgments between zero and non-alveolars. It is brought into a sharper focus by the fact that in their EPG analysis, they noted a 'marked, but counter-intuitive difference' in the EPG patterns for some of the zero alveolar patterns in comparison with the non-alveolar velar controls - for the former there was *less* lateral contact and a *more retracted* velar which results in less overall tongue-palate contact. They comment, that far from being an extreme form of assimilation:

> This is, in fact, residual evidence of a tongue body configuration appropriate for an alveolar: as the tongue tip moves up towards the alveolar ridge, the blade and pre-dorsum become concave, which reduces the amount of lateral contact in the pre-velar area. At the same time, this tongue shape will cause the velar contact itself to be more retracted. (p.54)

Gibbon et al (1993) presented evidence of a similar phenomenon in their cross-language study of EPG patterns in /kl/ clusters. For all speakers, the /k/ in the /kl/ sequence was more retracted than the singleton /k/. This contrast was interpreted as an inhibiting influence of the /l/ on the tongue body for the /k/ gesture.

Wright and Kerswill conclude that an incomplete 'merger' between full assimilations and non-alveolar, lexical velars is likely. But the question for them is: 'whether there is always a residual gesture, or just sometimes…does a gesture always result in an auditory percept?' (p.56) They point out that more sophisticated instrumentation, capable of showing more subtle tongue configurations, is vital to these issues. More recent studies of assimilation discussed in section 1.3.3, have utilised such instrumentation.

From the above review of studies of assimilation, the following summarising points can be made:

- A gradual loss of the alveolar stop is usually observed.
- Manner of alveolar articulation possibly determines assimilatory outcomes. There is some evidence that nasal alveolar stops are more susceptible to assimilation than non-nasal alveolar stops. Similarly in the case of /l/ vocalisation, phonetic parameters can determine the 'completeness' of the alveolar gesture. Before alveolar stops, full apical lateral closure is likely (e.g. 'build') whereas total vocalisation is likely in the context of back consonants (e.g. 'walk').
- The nature of the relationship between production of partial assimilations and perception has been questioned. There may or may not be a perceptible reflex of an incomplete alveolar articulation.
- Phonological models of post-lexical assimilation are inadequate.

This research can be seen in the context of a wider trend which has sought to attribute phenomena traditionally associated with the phonological level to phonetics. The recent 'Laboratory Phonology' paradigm:

> …can be interpreted as attempting to uncover experimentally whether critical phonological data is indeed phonological, or phonetic...It is unarguable that among the sound system alternations that languages evince are suppletive alternations and phonetic alternation - the problem for phonological theory is to predict that a given alternation is not phonological, but belongs instead to one of these other domains.
>
> (Scobbie et al, 1996)

Pierrehumbert and Talkin (1992) found that allophonic variants of /ʔ/ and /h/ are determined by low-level phonetic effects from the prosodic environment in which they appear. In a similar vein, Sproat and Fujimura (1993) proposed that the treatment of light and dark /l/ as categorically distinct phonological elements is redundant. They found that the position of /l/ within a syllable and its prosodic context alone determines whether it is realised as velarised or palatalised. The claim is that at no level in the psycholinguistic mapping, from abstract lexical representations to physical speech movement, is there a stage where the separate forms [l] and [ɫ] are distinguished for English. Zsiga's EPG work, discussed in more detail below, (1995) on the allophonic variation that can arise in realisations of /s/ to /j/, such as in *press your*, has shown that /s/ to /j/ assimilation is not simply categorical but is a gradient progression including intermediate 'compromise' articulations somewhere between [s] and [ʃ]. In common with Sproat and Fujimura (1993), Zsiga states that the quantitative variations she finds are due to the predictable and automatic properties of the speech production mechanism and not due to a postlexical palatalisation rule either at a phonological or phonetic level. Claims that speakers have limited stored information are similarly made by Browman and Goldstein (1990) who consider that there can be no mapping between the abstract and the physical in speech since gestures are phonological and phonetic entities. Phonological rules have no part to play because variation seen in the physical output of the gestural score is a result of variations in the temporal overlap between individual gestures and in magnitude of the gestures.

1.3.2 Assimilation and speech rate

Both Kerswill (1985) and Barry (1985) showed clearly using EPG that as speech rate increases so does the occurrence of assimilation. Kerswill observed this effect in data acquired from a single speaker and Barry from four speakers. Whilst the details of intra and inter-speaker variability are given in both studies the focus remains on wider issues such as phonological modelling of assimilation (Barry) and sociophonetic influences (Kerswill). Both studies categorise the data in the same way as Wright and Kerswill (1989) according to occurrence of full alveolar, partial alveolar and zero assimilation. Hardcastle's work on /kl/ coarticulation in different speech rates (1985) confirmed that by far the most robust influence on intergestural control of this sort was speech rate over and above the effect of prosodic/syntactic boundaries and vowel environment. VCV consonantal weakening has also been found to be a function of speech rate by Farnetani (1990) amongst others. She has shown using EPG that voiced stops in particular are reduced spatially by increase in speech rate.

The results of Kerswill's study is shown in Table 1.2 below. A single speaker was instructed to produce sentences which contain potential /t, d, n, nd/ assimilation sites and lexical control sequences such as:

boat covered	-	oak cupboard
cut price	-	cup prize
pen broke	-	Pembroke
hand gun	-	hang glider

Each sentence was read 20 times in each of the four speaking conditions shown below:

slowly and carefully
at a normal, comfortable speed
as fast as you can, but carefully at the same time
as fast as you possibly can

Table 1.2 distribution of articulation types across rate/styles for single speaker (taken from Kerswill (1985: 31))

		slow, careful	normal	fast, careful	fast
ARTIC. TYPE	complete assimilation	5	10	14	18
	partial assimilation	5	8	3	2
	absence of assimilation	10	2	3	0
	(SUM)	(20)	(20)	(20)	(20)
		STYLE			

Table 1.2 shows a clear increase in the number of assimilated tokens in the direction from the slower to the faster styles and Kerswill found this to be a statistically significant effect. There is, however, some evidence to suggest that when this subject was asked to speak quickly but clearly this tendency could be overridden. What is perhaps surprising is the relatively frequent occurrence of zero alveolars at the normal rate. Also surprising is that only half the tokens in the slow and careful rate were full alveolars. In common with Hardcastle (1994) it was found that the nasal alveolar (*pen broke)* and the nasal cluster (*hand gun)* are more assimilable than the non-nasal alveolars. Kerswill has two possible explanations, one phonetic and one phonological. Firstly assimilation is more likely because the formant transition cues leading into nasal consonants are less distinct than in the case of oral consonants (because of the 'acoustic obscuring effect' of the nasality in the preceding vowel). This means that it is harder to perceive place of articulation in nasals and is thus more ambiguous perceptually and more easily assimilated. The other explanation is that the homorganicity constraint on nasal + stop sequences in English, which means that an alveolar nasal must only be followed by another alveolar is:

> ...being extended to include cross-syllable environments as well. It may in fact be the case that the phonetic effects of nasality themselves gave rise to this constraint in the first place; whether or not this is so, we can envisage both influences acting together on this CSP [connected speech process].' (p.34)

The broader sociophonetic focus of Kerswill's investigation into CSPs (1985) is not shared by any other studies in this area. His hypothesis is that in the course of linguistic change, some connected speech processes become 'fossilised' as morphophonological rules, a situation which is one source of phonological change. That is, the process is maintained even after the factors which motivated it originally, such as phonetic environment, speech rate and articulatory care, have gone, for example in English [sj] has become [ʃ] in words such as *nation*. Crucially, these fossilised entities are always *discrete* in their application. Ohala (1993) has put forward a related argument that some sound patterns in languages are the result of fossilized coarticulation where synchronic/ phonetic contextual variation gradually became diachronic/ phonological variation via sound change. Kerswill and Ohala both consider that eventual fossilisation involves a change in the 'phonetic plan'. Kerswill also hypothesises that those processes which are currently *becoming* fossilised are more likely to be phonetically discrete and less likely to display degrees of partial assimilation. Judging by the data he presents, the alveolar to velar case is one which is not in the process of fossilisation. He notes that he has only referred to the extreme ends of the 'gradualness-discrete continuum'. Discreteness is straightforward but the description of gradualness is problematic on account of the lack of isomorphy between articulation, acoustics and perception[4]. In a similar sense, some articulatory gestures do not have acoustic or auditory consequences. Hardcastle and Morgan (1981) found, in a population of disordered speakers, the presence of intrusive articulatory gestures not detected auditorily, such as a velar EPG contact pattern during /ʃ/.

Kerswill proposes that phonetically gradual types of process are natural and inevitable since they are probably due to the nature of vocal tract motor control mechanisms like /t/ deletion in words like *mists* (Hewlett, 1981). This issue of the role of 'mechanical' effects in gradience becomes more important in the work of Holst and Nolan (1995) and Nolan (1992) discussed in upcoming sections. In order to confirm that some processes are undergoing change and are therefore less likely to be purely phonetically motivated, Kerswill measured a CSP's sensitivity to *speaking rate*, assuming that it will show less phonetic variation. The theory is that optional CSPs spread from casual speech to more formal or slower speaking styles until they become obligatory. While Kerswill identifies those processes fossilised into morphophonological rules as separate to those with a synchronic phonetic motivation, he points out that they are both points on the same continuum. To summarise, fossilised forms are always discrete, changing forms are variable but tending towards discreteness and 'natural' coarticulatory forms are always gradual.

With regards to social variation, Kerswill found that 'the working class vernacular of Durham contains a number of CSPs differentiating it from RP, while at the same time

[4] That is, a change in one of these might not lead to a comparable change in another. A listener can hear a series of consonant-like stimuli and will assign them to a particular phoneme of his or her language. If the

lacking certain CSPs which are found in RP' (p.19). The example is given of deletion of the final vowel in 'into', in the phrase 'into the car' or regressive voicing assimilation resulting in realisation of *like me* [laɪg meɪ]. Conversely , alveolar place assimilation such as [ðaʔk gɜ:l] *that girl* is reported to be rare in Durham English vernacular.

Like Kerswill, Barry (1985) showed that rate has a considerable effect on frequency of assimilation but does not actually determine assimilation. Barry examined alveolar to velar sequences produced by three speakers of RP or near-RP English. Each subject read sentences containing sequences of the type *oak cupboard/ boat covered; late calls/ make calls; cut price/ cup prize*: twice at normal conversational speed and twice as quickly as possible.

The pooled results for all speakers showed a predictable increase in complete assimilation from the normal condition to the fast as possible condition, as Kerswill had found. It is important to note that these results are for alveolars preceding velars *and* bilabials. Results for all 3 speakers combined is given in Table 1.3 below:

Table 1.3 distribution of EPG assimilation types – results for 3 speakers combined (from Barry, 1985)

	full alveolar	partial assimilation	complete assimilation
casual	23	14	11
fast	15	15	18

Barry also notes some inter-speaker preferences. Subject 1 tended to always assimilate, subject 2 tended to assimilated partially and subject 3 preferred not to assimilate. Scrutiny of the results shows that there is a very nearly identical occurrence of partial assimilations in conversational *and* fast speech. Interestingly, Barry found that alveolar to bilabial sequences appear to yield far fewer residual articulations than alveolar to velar sequences. Table 1.4 shows the scores for the different sequences. Scores for partials are out of a total of 12 repetitions per experimental sequence.

stimuli differed between themselves acoustically the listener would still respond to them as members of a single phoneme category.

Table 1.4 number of partial assimilations out of a possible 12 (casual and fast speech combined) for 3 speakers (data from Barry, 1985)

alveolar-bilabial	**partial assims**	**alveolar-velar**	**partial assims**
man made	1	*hand grenade*	4
cut price	0	*foot came*	3
		boat covered	5
		late call	6
		can't come	4
		can come	6

An interesting situation arises whereby sequences which involve different articulatory subsystems, here the tongue and the lips *man made* and different parts of the same organ *can come*, are more likely to be completely assimilated.

One final aspect of Barry's study which deserves attention is the fact that one of the speakers produces very few residual articulations compared to the other subjects. Of the 32 tokens yielded overall from this subject only 2 partials were produced. The other subjects produced 18 and 9 of these respectively, again out of a possible 32. Therefore it seems that some speakers are more 'tolerant' of partial forms than others.

Nolan (1992) has pursued the question posed by Local (1992) quoted at the beginning of this chapter which refers to the situation in assimilation where one thing accommodates to another: '*what* are these things?...*where* are these things?' In particular Nolan focuses on the 'where' of assimilation by considering on the issue of phonological modelling. A featural model (see Figure 1.2) is quickly dismissed on the grounds that 'such a notation fails to show why certain subsets of features, and not other subsets, seem to operate in unison in such assimilations, thus...failing to capture the traditional insight that these changes involve assimilation of place of articulation.' (p.262) But the shortcomings of autosegmental phonology, where features are organised into functional groupings, are similarly serious in view of the graded nature of alveolar to velar assimilation, namely the fact that representations are categorical. Intermediate alveolar to velar forms would be modelled as (b) in Figure 1.12 where the node for C_1 can be associated to the place node of C_2 without losing its own place features. Taken a step further complete place assimilation occurs (c).

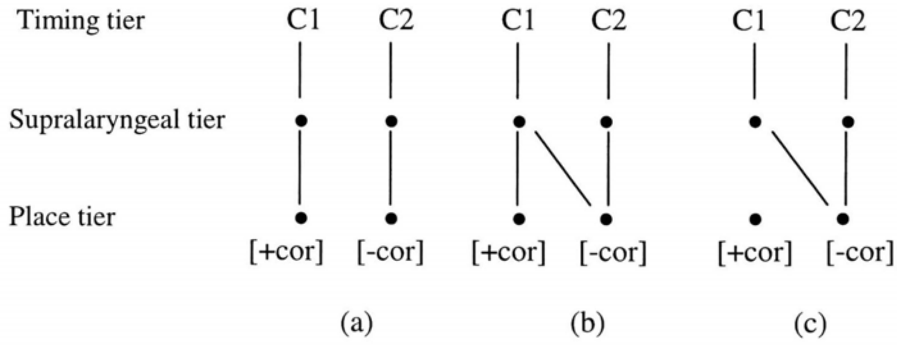


Figure 1.12 autosegmental models of alveolar to velar place assimilation (Nolan, 1992)

Nolan comments that there is no reason other than goodwill why this interpretation should be made from this notation. The main problem lies in form (b), the phonological 'complex segment'. Technically, this form represents a double articulation because both the coronal and velar entities have equal 'strength'. But in the case of residual alveolars, the coronal is not only variable but contrasts with the velar in that a velar is resistant to assimilation. Delinking as in (c) is not appropriate either since there is no coronal specification at all. In rejecting the idea that assimilation is a phonological process, Nolan then turns to the level of phonetic implementation where these effects maybe arise because of physiological constraints 'and inherent characteristics of its 'programming' as proposed in the theory of 'coproduction' (Fowler, 1983). However, accounts based solely on the vocal mechanism soon fall down when it is considered that place assimilation is not universal (see section 1.4). Evidence that not all languages display assimilation inevitably means that:

> it is a phenomenon over which speakers have control. This will provide further evidence that a greater amount of phonetic detail is specified in the speaker's phonetic representation or phonetic plan than is often assumed. (p.278)

In reply to the issues raised by Nolan, Hayes (1992: 'commentary', same volume) defends phonological modelling of assimilation. In particular, Hayes responds to the idea that phonology cannot capture gradient assimilation, as in the case of alveolar to velar sequences, because this is not phonology's domain. It 'should not contain quantitative information. The proper level at which to describe variability of [alveolar] closure is actually the phonetic level.' (p.282)

Hayes proposes that there should be two distinct processes to explain gradual alveolar to velar assimilation. There should be a phonological spreading rule, Place Assimilation and a phonetic rule, Alveolar Weakening, which act in combination during the derivation. Hayes stipulates that the Place Assimilation rule produces not an entirely delinked place node for the coronal but a 'corono-dorsal complex segment'. This complex segment can be seen in Figure 1.12 (b) above. Once this underlying complex segment is in place Hayes says 'we have to provide a way of varying the degree of closure made by the tongue blade. Since phonological representations are discrete rather than quantitative, they are inappropriate for carrying out this task…The rule responsible for weakening alveolar closure is a phonetic rule, and as such it manipulates quantitative values.' (p.283) This rule is formalised as:

> *Alveolar Weakening*
>
> Depending on rate and casualness of speech, lessen the degree of closure for a COR autosegment, if it is [-continuant] and syllable-final.

So, from the phonological output (the corono-dorsal complex segment), degree-of-closure targets are assigned to the abstract specifications CORONAL and DORSAL by the phonetic component whereupon the Alveolar Weakening rule can be applied to lessen the degree of closure for the CORONAL target. Notice that before the weakening rule is applied, the phonological information must be brought one step closer to phonetic realisation by the degree-of-closure process. Figure 1.13 is Hayes' illustration of how these separate processes combine:

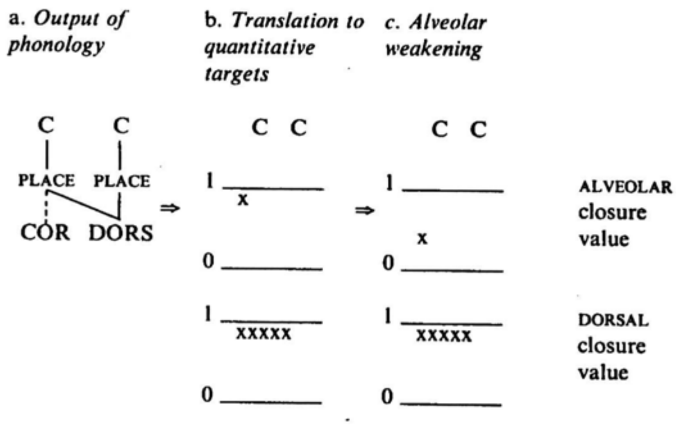


Figure 1.13 stages of derivation for alveolar assimilation: phonological assimilation rule and phonetic weakening rule (from Hayes, 1992: 284)

Hayes also goes on to argue that having two distinct processes means that in principle, one rule might apply in the absence of the other. He gives the example of the articulation of /t/ which is often weakened even when no other segment follows, although this can only occur when it is 'covered' by a simultaneous glottal stop. If Hayes asserts that the phonetic rule can be applied independently of the phonological rule, then the alternative situation should be equally valid, that the Place Assimilation rule can be independent of phonetic more-or-less reduction. The problem here is that the output of the Place Assimilation rule, a complex segment, specifies that at some level in the generation and execution of the sequence an alveolar is always present. It seems that this complex segment, as the phonological component of the whole process, is only proposed at all because it subsequently allows an alveolar weakening rule to be applied, i.e. it has the preserved alveolar specification to work on, and because Hayes is a phonologist and wants to solve the problem using phonological analysis. When place assimilation is referred to by Nolan as a 'gradual process' it should be the phonetic weakening component only which is regarded as gradual, Hayes argues, and not the place assimilation rule proper. Hayes also argues that the Place Assimilation rule should still be regarded as discrete regardless of the data.

1.3.3 Assimilation and combined EPG/EMA methodology

So far in this review of EPG studies of place assimilation there has been an underlying methodological problem which Nolan (1992: 267) has summarised:

> Because it [EPG] records only tongue contact, it gives no information on possible differences in tongue position or contour short of contact. This means, in particular, that it is premature at this stage to conclude that even zero alveolar tokens represent a complete assimilation of the tongue gesture of the first consonant to that of the second.

Kühnert's work (1993) on alveolar (/t, d/) to velar assimilation forms an important contribution to the literature and uses a technique in combination with EPG which can track lingual movements in 2-dimensions. Apart from some x-ray micro-beam data acquired by Browman and Goldstein (1990), little is really known about the overall shape and configuration of the tongue during the execution of (assimilated) alveolar to velar sequences. This combined EPG/EMA study looked systematically, for the first time, at the possible occurrence of the type of residual alveolar gesture which is undetectable using EPG, that is, where the tongue tip/blade is raised only partially leaving no trace on the tongue-palate contact patterns. In common with other studies, Kühnert is fundamentally concerned with the extent to which assimilation can be attributed to articulatory reorganisation or gestural reduction. A principal aim of the study is to assess the adequacy of Articulatory Phonology, a theory which dispenses with the need for phonological rules of reorganisation and which states that careful speech and casual speech realisations derive from the same structure which forms the input to the gestural score. On the basis of her data she concludes that gestural phonology has some points in its favour but basically cannot account for cases where there is a complete loss of the coronal gesture in alveolar to velar sequences. She also questions the validity of the mechanism which is said to govern reduction in magnitude of gestures. The study is also cross-linguistic with data from English and German speakers and it looks in detail at fast speech effects. Highlighted in particular is the non-correspondance between apparent EPG-defined complete alveolar assimilation and EMA-defined tongue configurations.

Kühnert took 3 German and 2 English speakers who produced sentences containing the set of sequences /t#k, k#k/ and /t#h/ or /d#k, g#k/ and /d#h/. Due to word-final devoicing in German, the German speakers produced only the former voiceless set and in order to avoid the complicating factor of frequent glottalisation that may occur as an accompaniment to or in place of /t/ in English, the English speakers produced only the voiced alveolar sequences. The /t#h, d#h/ items were included to establish that alveolar assimilation is not just a result of syllable-final weakening, since the glottal fricative is not known to motivate assimilation. In total, each subject produced 30 repetitions of each sequence at normal speech rate and 30 at as-fast-as-possible rate (a much larger number of tokens than previously elicited by other studies). Stimuli was devised to include a number of different combinations of vowels thus: German: V_1=/a, ɛ, ɪ, ɔ, ʊ/ and V_2=/ɑː, a, ʊ/; English: V_1=/ʌ, e, ɪ, ɒ, ʊ/ and V_2=/ɑː, ʌ, ʊ/. Examples of test utterances are shown below:

German: /ɔC_1#C_2ɑː/	English: /ɒC_1#C_2ɑː/
Dein Spottt kam überraschend.	The rod can't crack.
Der Schock kam nach dem Unfall.	The dog can't bite.
Sie will Kompott haben.	It is an odd harem.

Her results showed speaker-specific variation in assimilatory behaviour, although no language-specific effects. For all subjects assimilations occurred almost exclusively in the fast speech environment. 2 subjects (one German 'GER1' and one English 'ENG1') assimilated in about half of the 30 utterances and in such a way that the tongue tip raising gesture was most frequently present. 2 subjects (GER2 and ENG2) hardly ever assimilated and when they did only very little: 'In the latter cases upward tongue tip movements were always easily detectable in the EMA displacements as well as in the EPG contacts.' (p.268) Unfortunately the full EPG patterns are not shown. The remaining subject GER3, assimilated in almost all utterances 'and, with very few exceptions in which a slight residual coronal was still distinguishable, all assimilations were so extreme that it seemed as if a complete gestural deletion process was at work.' Regarding the possibility that assimilation is due to syllable-final weakening (as suggested by Hayes, 1992) and thus that the loss of an alveolar might occur regardless of what sound followed, Kühnert reports that this is not the case since, even for the 'dramatically assimilating speaker GER3', the coronal target before /h/ was preserved.

Kühnert presents two main strands of interest arising from the results which will be discussed in turn below. The first regards her assertion that assimilation is a gradual process for *all* speakers and, less importantly, that it is gradual across speakers, i.e. speakers seem to prefer a certain range within the continuum tokens showing full alveolar closure to apparent complete gestural deletion. The second regards the relation between EMA and EPG data.

Whilst she has shown that subjects differ in the extent of the variability in their assimilatory behaviour, she places overriding significance on the fact that all subjects 'execute residual coronal movements at least some of the time' (p.271) and she interprets this as a point in favour of the gestural blending perspective on assimilation. Although she asserts that gradience is the norm in the speakers' productions of /t#k/ or /d#k/ clusters, it seems from her results that it is more likely that it is more the norm for some speakers than for others. Figure 1.14 below is Kühnert's summary of EMA tongue tip and tongue dorsum positions at the beginning of the experimental sequences for subjects GER1 (who assimilates in about half the utterances) and GER3 (who hardly ever assimilates and if so only very slightly). The aim of these displays is to compare the position ellipses of the different lexical forms for each speaker to detect incomplete assimilations in the form of reduced tongue tip raising gestures associated with the alveolar stop.

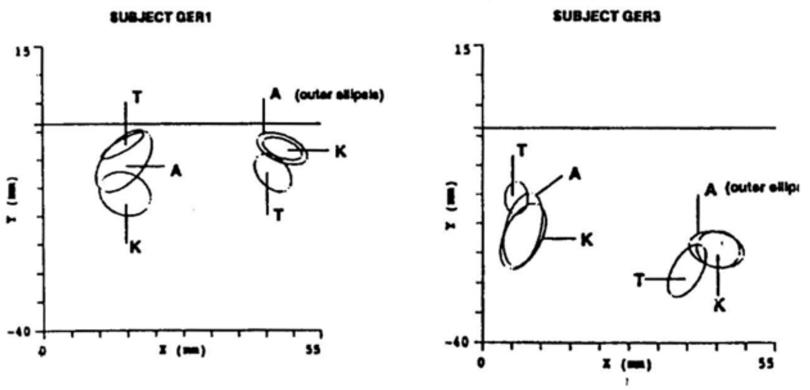


Figure 1.14 articulatory positions of tongue tip (left-hand side of graphs) and tongue back (right-hand side) at the beginning of consonant sequence for fully produced alveolars (T), assimilated alveolars (A) and velar controls (K) for speakers GER1 and GER2 (from Kühnert, 1993: 269)

Both subjects GER1 and GER3 are said to show gradient assimilations. It is pointed out that the tongue tip ellipsis for the assimilated tokens 'A' produced by subject GER1 is more variable than that for the non-assimilated tokens 'T' or the velar controls 'K' between which the assimilated tokens ellipsis lies. But for subject GER3 'in contrast, the ellipses of A and K tokens are essentially the same. Only in a few assimilated productions does the tongue tip extend further upwards than in the velar control sequences'. (p.269) We must assume that these exceptions are located in Figure 1.14 within the area where the 'A' ellipsis extends vertically beyond the area of overlap with the 'K' ellipsis. These exceptions were previously described by Kühnert as having 'a slight residual coronal' but it is difficult to see how these could be described as slight since they would have to involve considerable vertical tongue tip displacement close to that for the achievement of full closure. Kühnert contrasts the situation for GER1 and GER3 and yet concludes that all subjects are the same in that they all produce a continuum of remaining coronal gestures.

The difficulty lies in Kühnert's use of the term continuum. GER1's productions uncontroversially form a continuum. The 'A' and 'K' ellipses overlap but have essentially different patterning. In GER3's case, however, the 'A' ellipsis does not fall between the other two but overlaps to such an extent that it seems difficult to make a case for them being different. If the 'A' ellipsis was removed from the display altogether, then the full alveolar ellipsis 'T' and the velar control ellipsis 'K' would form a continuum on their own because the edges of these latter ellipses make contact. Since both the 'A' and 'K' tongue tip ellipses are so variable vertically, it is hard to see how tokens that reach a certain height, which here are those assimilated alveolar tokens which only just fail to make contact with the alveolar ridge, should be regarded as a different phonetic form i.e. a residual alveolar. Perhaps GER1 and GER3 are more different from each other than they are alike, as Kühnert herself has argued on one level, and in that case it is the differences in the variability of tongue tip height for the velar controls which is the determining factor.

Kühnert's second main theme regards the relation between EMA and EPG data. Of principal interest is the question of whether identical EPG patterns of assimilated alveolar-velar tokens and velar-velar control tokens reveal different overall tongue configurations. That is, will different lexical forms always result in different articulatory outputs? She explains:

> Unfortunately, the answer is that they may do but that they need not do so. Detailed examination of the two data categories reveals that more or less identical EPG patterns stemming from different lexical forms can either have quite different underlying overall tongue configurations or (apparently) identical ones. p.267

Figure 1.15 shows Kühnert's examples from the data for GER1. Trajectories of tongue tip (left), tongue mid (middle) and tongue back (right) are shown from the middle of the preceding vowel to the middle of the following vowel during 'normal' production of /atkɑ:/ (top left) an assimilated /atkɑ:/ (top right) and an /akkɑ:/ (bottom). The line which joins all three coil trajectories indicates the overall tongue configuration at the point of maximal tongue tip displacement for the top two panels and at the beginning of velar closure in the bottom panel. Because there is no perceivable tongue tip gesture in the /akkɑ:/ production it makes sense to indicate tongue shape at the onset of velar closure. The EPG frames that appear below each display represent the same point in time as the solid line is taken from.

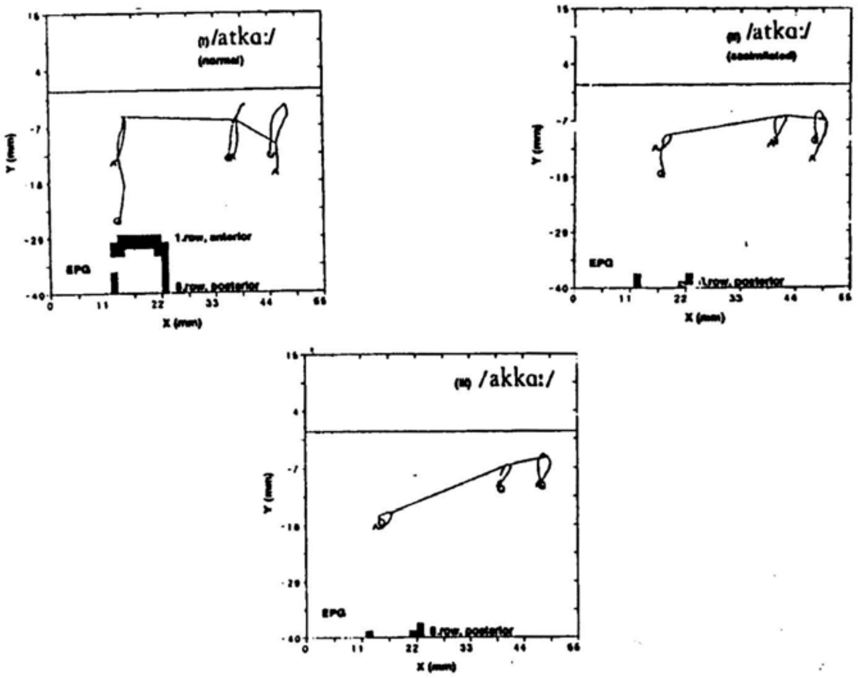


Figure 1.15 trajectories of tongue tip (left), tongue mid (middle) and tongue back (right) from the middle of the preceding vowel to the middle of the following vowel during a normal production of /atkɑ:/ *(top left panel); an assimilated production of* /atkɑ:/ *(top right panel) and a production of* /akkɑ:/ *(bottom panel). The solid line indicates the overall tongue configuration at the moment of maximal tongue tip displacement in the top two panels and at the beginning of velar closure in the bottom panel. These moments in time are also represented by the individual EPG frames at the bottom of each graph. (from Kühnert 1993: 267)*

What this display shows is that in contrast to the 'normal' unassimilated /atka:/ production, the assimilated form in fast speech shows no contact on the alveolar ridge resulting in an EPG contact pattern very similar to that of the fast velar control token at the bottom. But there is a coronal gesture, somewhat reduced in magnitude which is overlapped by the velar unlike the coronal gesture in the unassimilated version. Importantly, examples such as this where there is a discrepancy between EPG and EMA data tended to occur in the low vowel environments /a, ɔ/ for German and /ɒ, ʌ/ for English. 'Thus only when the overall articulatory setting allows for it, in the form of a preceding low tongue configuration, can a residual raising movement of the tongue tip still be executed without affecting the palate.'(p.268) Wright and Kerswill (1989) speculated that retracted velar assimilatory patterns for *beg/bed, leg/lead* and *dig/did* might be a consequence of this particular tongue configuration although these sequences have different vowel contexts compared to Kühnert's (see section 1.3.1). Kühnert surmises that 'it might sometimes be unreliable to classify gradual assimilations on EPG evidence alone.' But a major drawback of this study is the fact that the full EPG patterns produced by subjects are not provided and so there is no opportunity to see if intermediate assimilatory forms are actually visible on the palate. Furthermore, alveolar nasal stops may show different assimilatory effects.

She also notes instances where indistinguishable EPG patterns resulting from lexically different forms also have *identical* underlying tongue shapes and thus, she says assimilation is discrete. For subject GER3 who nearly always assimilates, this is the case for about 85% of his tokens. It is these cases which lead her to propose that gestural phonology cannot account for all assimilatory forms through the sliding in time of gestures on separate oral tiers alone. It is not obvious what determines which coronals undergo weakening (reduction in magnitude is also considered gestural blending) and which do not. There must, she concludes, be an independent phonetic motivation which causes the variation in magnitude of the alveolar gesture.

1.3.4 Mechanical factors in the production of alveolar to velar sequences

In an EPG study of articulatory timing in consonant sequences, Hardcastle and Roach (1977) found that front-back /tk/ ($VC_1(\#)C_2V$) sequences have shorter C_1-C_2 times compared to back-front sequences /kt/. These consonant sequences are interesting because many demands are concurrently placed on a single articulator, in these cases, the tongue. Byrd and Tan (1996) also note that the timing of front-back sequences is different to the timing of back-front sequences but expresses the difference in terms of degree of overlap, front-back sequences are more overlapped on the basis of EPG data. Hardcastle and Roach have a physiological account of why this discrepancy in timing occurs. They say that in the case of /tk/, sensory information, as the tongue makes contact along both sides of the palate and alveolar ridge for C_1, causes a contraction of the intrinsic muscle, the inferior longitudinalis for the C_2 velar closure. But for /kt/, the movement towards /t/ which is triggered as the back of the tongue makes contact with the velar region, probably, they say, involves the genioglossus muscle which is an extrinsic muscle *and* the superior longditudinalis. They comment: ‘Because it [the alveolar stop closure] involves repositioning the tongue body by the slower-contracting extrinsic muscles the movement towards the /t/ is relatively delayed.’ (p.34) Coronals are thus slower to produce than velars. This finding has a relevance to the phenomenon of assimilation and what is thought to cause it. Byrd and Tan (1996) question this gesture execution hypothesis and asks whether there would really be a difference in the time taken to deploy one intrinsic muscle compared to one extrinsic muscle: ‘the well-supported conception of muscle groups organised into coordinative structures suggests that it is unlikely that the exact number of muscles involved should create differences in the time between articulations.' (p.234) Byrd instead proposes that the differences in production of front-back and back-front sequences is due to interactions between perception and production whereby speakers will make less of an effort to preserve the less salient, unreleased [d] perceptual cues. Since the endings of words are less important perceptually to listeners compared to the beginnings of words, speakers are able to distribute articulatory effort accordingly.

Figure 1.16 below shows the EPG 'contact totals profiles' for 3 of her 5 speakers' multiple productions of /d#g/ and /g#d/. These profiles were generated from taking the percentage of contacted electrodes in each articulatory region (front or back) across time. Speakers were instructed to produce the stimuli, embedded in words, at a fluent reading rate. For each speaker there is a pair of displays, the left hand display shows contact for /d#g/ and the right hand displays show contact for /g#d/. The vertical axis shows percentage contact and the horizontal axis represents time in EPG frames (10ms sampling rate). The solid line curves show contact for the alveolar stop (front) and the dashed line curves are for the velar stop (back).

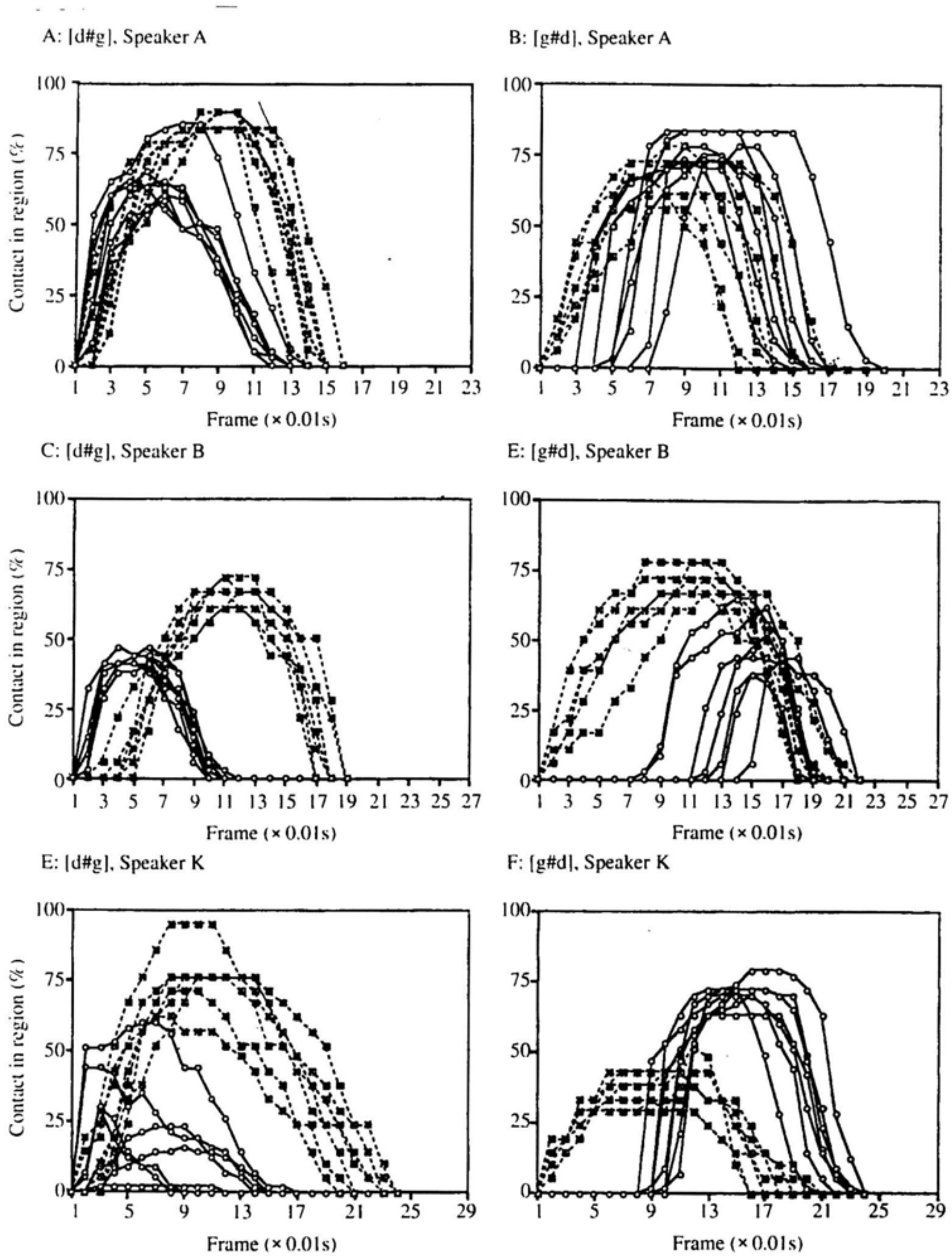


Figure 1.16 Contact profiles for /d#g/ (left) and /g#d/ (right) for three speakers ---o--- *corresponds to contacts in front region* ···•··· *corresponds to contacts in back region (from Byrd, 1996: 218)*

These contact 'profiles' or 'totals' give an indication of the overlap that takes place between the front and back consonant and they also show how 'weakened' the alveolar stop is. Shallower alveolar stop curves indicate that full stop closure may not be achieved and may even be completely absent.

For speaker A and K the onset of the /d#g/ stops do appear to be simultaneous (in terms of tongue–palate contact) whereas for speaker B there is a clear tendency for velar contact to start slightly after contact for the alveolar stop has been established. However, for speaker K there is much more variation in amount of contact achieved for velar closure, and more significantly, for alveolar closure. It is important to note that for this speaker, the amount of alveolar stop closure forms something of a continuum while some flat curves show that there is no alveolar closure produced at all. By comparison this same speaker's C_2 alveolar curves for the /g#d/ sequence do not show so much intra-repetition variation in amount of contact.

Byrd's other focus in this work apart from the influence of sequence timing on consonantal place of articulation is the influence of consonant manner and syllable structure on timing. With regard to syllable position, she found that onset clusters (#CC) are less overlapped and less variable in timing than coda clusters (CC#) and heterosyllabic clusters (C#C). This highlights the importance of syllable initial acoustic information over syllable final information. With regard to consonant manner, she found stops to be more overlapped by a another following stop than fricatives, commenting that the 'degree of overlap permitted in a particular sequence may be a function of the constriction degree of the consonants involved, with less overlap expected in sequences including a fricative, as a coarticulated closure would impede the airflow necessary for frication.'(p.229)

Observation of the production of two sequential stop consonants has shown that there is normally a short period of overlapping or simultaneous closure (Ladefoged, 1993). Furthermore, in most cases, the closure for the first stop will not be released until the closure for the second has been formed. Hardcastle and Roach (1977) looked at 3 subjects' productions of a variety of sequences: /kt/, /tk/, /kp/, /pk/, /tp/, /pt/ (4 repetitions each). They found that in only 32 cases out of 372 was the first stop released before the onset of the second closure, indicating the high level of articulatory cohesiveness involved.

Marchal (1988) and Recasens et al (1993) were both interested in the possibility that stop sequences of the type VC_1C_2V are not executed in a linear order but as a more cohesive unit. Their work argues against the basic theory of coarticulation. Kent and Minifie (1977: 115) state that 'coarticulation is a conceptualization of speech behaviour that implies discrete and invariant units serving as input to the system of motor control, and an eventual obscuration of the boundaries between units at the articulatory or acoustic levels'. By contrast and on the basis of EPG analysis of /tk/ and /kt/ produced by two French speakers, Marchal hypothesises:

> The consonantal mechanism initially produces a rather neutral consonantal gesture: the whole lingual body moves upward, and one cannot assess at first whether or not it is preparing for the articulation of C_1 or C_2. The tongue first seems to adopt a stable position where it can presumably obtain tactile feedback about its current location and articulatory target. It can then differentiate the upcoming articulations, and produces C_1. (p.293)

Recasens at al (1993) take a similar coproduction view of stop sequences concluding that they 'are planned to a certain extent as a homogenous production event'. They arrived at this conclusion by measuring the location of coarticulatory effects from a given consonant on an adjacent consonant. They were also interested in the production of /tk/ and /kt/ sequences and wanted to know 'whether coproduction between consonants articulated with contiguous lingual regions in the clusters /tk/ and /kt/ results in mere gestural overlap or into some perturbation of the articulatory configurations for /t/ and /k/ i.e. gestural blending' (p.338) Experimental items included all combinations of the vowels /i/ and /a/. Data from 3 speakers were collected, one was a speaker of Catalan and the others were American English speakers. In comparison to other stop sequences (/tp/, /pt/, /pk/ and /kp/), the sequences /tk/ and /kt/ showed additional contact at the front and the back of the palate. They found that the front closure in these clusters is larger and more retracted than for singleton /t/. They conclude from this that

> Coproduction between /t/ and /k/ in the sequences /tk/ and /kt/ leads to intergestural blending about the cluster midpoint. Blending affects a tongue front gesture and a tongue dorsum gesture, and involves a shift in primary articulator and in place of articulation with respect to the two lingual consonants alone. The articulatory outcome of this blending process is comparable to a single gesture...The case presented in this paper indicates that blending involving the same articulatory structure (i.e. the tongue) may take place between two semi-independent articulators. (p.351)

Unfortunately, no clear indication is given of whether this blending occurs for sequences in all vowel contexts or whether this applies only to sequences which precede the high-front vowel.

Observations such as these are said to reflect simultaneous organisation of the stop consonants at the production level. Recasens et al (1993) cite the work of Catford (1977) who states that the duration of overlap between C_1 and C_2 is between 30-40% of the combined duration of the two consonants. Recasens et al comments: 'This hypothesis is highly compatible with the notion of progressive articulatory adaptation in time within the cluster', clearly an approach which strongly contrasts with the idea of a homogenous production event.

1.3.5 'Same articulator' assimilation

Holst and Nolan's acoustic study of [s] to [ʃ] assimilation (1995) asks 'is segmental accommodation all in the mouth or is some of it in the mind?' That is, when adjacent segments have different specifications for the same articulator is the result complete assimilation (motivated by a high-order rule) or an articulatory 'compromise' gesture. The latter is predicted by Browman and Goldstein (1989) where both gestures would occupy the same oral tier on the gestural score. For instance, in the case of /t θ/ the predicted output would be a blended segment:

> ...the observed motion of the TT [tongue tip] tract variables resulting from overlap and blending should differ from the motion exhibited by either of the individual gestures alone. In particular, the location of the constriction should not be identical to that of either an alveolar or a dental, but rather should fall somewhere in between. (p.219)

Holst and Nolan's materials involved sentences designed to capture the potential site of assimilation [s] to [ʃ] as in *Before a shop assistant restocks shelves all old produce must be removed* and acoustic, durational and spectral analysis of the fricatives was carried out. In fact the sentences were also designed to capture the [s] to [ʃ] sequence with the intervention of a clause boundary. as in *Before a shop assistant restocks, shelves ought to be at least half empty.* The 12 subjects were instructed to read each sentence aloud at a comfortable speed as fluently as possible. Comparison of effects at different speaking rates was not of experimental interest here.

What they found was that [s] and [ʃ] across a boundary yields a continuum of forms from partial loss to apparent total loss of [s]. They found four types of outcome and these are illustrated in Figure 1.7 below: the first has two unambiguous stable patterns of acoustic energy corresponding the [s] and [ʃ], which they call Type A. The second and third are 'contour segments' as predicted by Browman and Goldstein which they call Types B & C. These show patterns of energy on spectrograms which glide between an [s]-like form and a more [ʃ]-like form. These contour segments can be further subdivided, it was discovered, between those which have a discrete period of static [s]-like friction at the start of the entire friction block and those where the friction block describes one long transition towards [ʃ]. These types are viewed as the 'kind of mechanical dynamic output of a device required to reach two incompatible targets in a short time.' The third is where [s] has become phonetically identical to [ʃ] which they call type D. Spectrally, this type is defined by a single stable period of [ʃ]-like friction. Thus there is one type of assimilation that is gradient and one that is 'all-or-nothing' categorical. Figure 1.17 shows Holst and Nolan's illustrations of these types. Another finding was that complete assimilation of the sequence (type D), but not gestural overlap types, is blocked across a clause boundary.

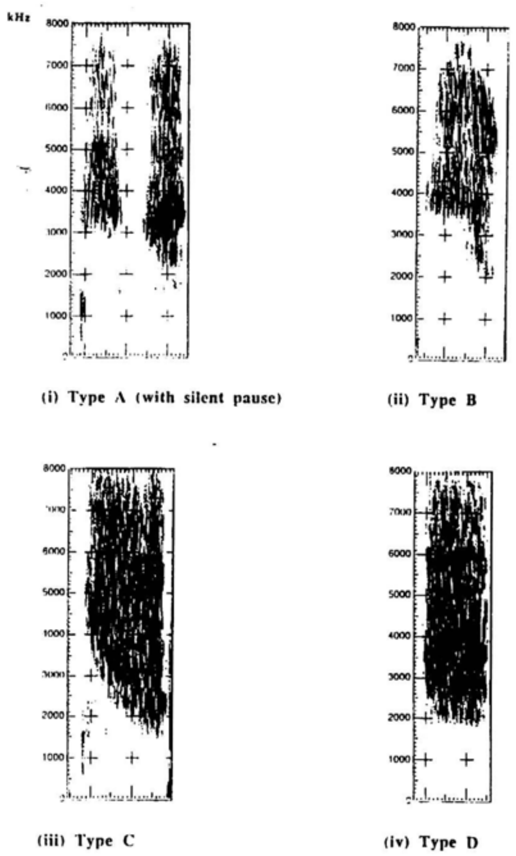


Figure 1.17 Spectrograms (0-8 kHz) of four [s] to [ʃ] *assimilation types (from Holst and Nolan, 1995: 322)*

According to Holst and Nolan, type D tokens present a form outwith the explanatory framework of Articulatory Phonology by virtue of the fact that the [ʃ] yielded by assimilation, is *longer* than canonical [ʃ] and this difference is highly statistically significant. This durational distinction cannot be accounted for by simple gestural overlap: 'The extra duration of [ʃ]-like friction...must mean that some duration "belonging" as it were to [s] has been articulated as [ʃ] - that is, in a very real sense, a process of assimilation has occurred.'(p.328) This process, it is argued, can only be modelled as a phonological assimilation rule. The other types B and C, which are articulatory 'hybrid' contour forms between [s] and [ʃ], are considered adequately handled by Articulatory Phonology as the result of competing gestures. On the subject of type D forms, Nolan, Holst and Kühnert (1996) comment:

> the first [s] gesture ...had been replaced by an [ʃ] gesture before the phonological representation was given its articulatory representation - involving, of course, considerable overlapping of the (now) two [ʃ] gestures bringing the duration of the fricative event down to near, but still greater than, that of the singleton [ʃ]. (p.116)

The primary contribution of this research is the idea that as phonetic weakening/ blending reaches a certain point, a threshold is reached whereupon a phonological assimilation rule is applied (type D). It can only be assumed that this occurs for all speakers.

In terms of phonological modelling, type D is the equivalent of the delinking shown in Figure 1.12 (c). Complex segments shown in Figure 1.12 (b) are acceptable for types B and C but are considered surplus to requirements since this situation arises naturally anyway from the universal workings of the articulatory mechanism:

> The crucial fact is that the speaker does not need to *do* anything to produce types B and C - they arise naturally from the characteristics of the articulators and their targets. There is therefore no motivation to invoke a specific cognitive operation, that is, a phonological rule, to model them. The kind of phonological configuration suggested for "contour segments" is therefore inappropriate and unnecessary. (Holst and Nolan, 1995: 329)

Complex segments are also considered inappropriate by Nolan (1992) for modelling alveolar to velar partial assimilation on the grounds that the predicted output of double articulations is wrong.

In defence of Articulatory Phonology, Catherine Browman comments on Holst and Nolan's results (same volume, 1992). She takes up her objection from their acceptance of phonological delinking Type D articulations. Browman would view what for Holst and Nolan is the product of a phonological assimilation rule as 'falling in a continuum with types A-C, so that type D would also be a type of increased gestural overlap between two gestures, with more overlap than in types B and C, but no complete overlap (because the duration is greater than that for [ʃ]).' (p.335) Furthermore, Browman questions the assertion that type D is the result of a re-specification process when their only evidence is acoustic. While there is a lawful relationship between articulation and the acoustic signal the relation is not always linear.

> There may be many positions of the tongue that correspond to a single set of acoustic values for, say, [ʃ]. Therefore, it is not possible to know from the acoustic signal exactly what the articulation is. That means it is also not possible to use the acoustic signal to determine whether there is one underlying feature specification, or two nearly completely overlapping feature specifications. p.336

Zsiga (1995) has provided EPG and acoustic evidence that while lexical /ʃ/, whether underlying as in *fish* or derived as in *impression* (*impress →impression*)[1] is categorical, postlexical /ʃ/, brought about from the production of /s/ to /j/ in for example, *miss you, press your* is the result of a gradient process. Featural representation is said to best handle the lexical /ʃ/ phenomenon and Browman and Goldstein's gestural model handles the post-lexical phenomenon. Zsiga says there is no place for phonological processes in the case of /s/ to /j/ coarticulation but she states that this gradient behaviour arises automatically from the properties of the speech production system. This sort of coarticulation caused by the inherent sluggishness of the tongue, is distinguished from coarticulation which may be governed by language-specific principles and rules but which still deals with non-categorical more-or-less variations. In a study of Igbo post-lexical phenomena, Zsiga (1997) found that Igbo vowel harmony is categorical but Igbo vowel assimilation is gradient. Again, she proposes that phonological rules are unnecessary for gradient processes because this arises naturally from the properties of the speech production apparatus. In common with Holst and Nolan (1995), Zsiga finds categorical and gradient processes occurring for the same articulatory sequence. There is no true comparison, though, since alternation pairs like *confess* and *confession* are Zsiga's categorical forms. While /s/ to /j/ has been lexicalised in *confession* she still considers this a 'process' handled by active phonological rules. Holst and Nolan's position is more radical because categorical and gradient forms are said to exist within a single post-lexical process.

1.3.6 Bilabial to velar place assimilation

Bilabial to velar place assimilation (*pk*) in Korean and English has been investigated by Jun (1996): 'Place assimilation is not the result of gestural overlap: evidence from Korean and English'. The purpose of this articulatory and perceptual study was to gauge the distinct roles of gestural overlap and gestural reduction for *pk* sequences within words. The author assumes that in cases of *pk* (e.g. /np +ko/ → [nukko]) in Korean a perceived deletion has occurred, but what cannot be assumed is whether it is gestural reduction or gestural overlap

[1] Zsiga assumes that 'the regular alternations relating words such as *confess* and *confession* should be expressed in the grammar as phonological rules' p.293

which is the source of this. More specifically, Jun challenges Browman and Goldstein's view that phenomena like place assimilation is mainly due to overlap. In particular, he argues that the perception of place assimilation requires more than overlap alone:

> The present study shows that at least some cases of casual speech place assimilation are not susceptible to an overlap-based account; gestural overlap alone cannot give rise to perceptual place assimilation. Instead, gestural reduction is argued to play the decisive role in casual speech place assimilation. Further, it is argued that gestural reduction is speaker-controlled; it does not result directly from physical constraints on speech-production mechanisms. (p.378)

Jun's experimental approach to this was unusual in that he acquired aerodynamic data. Oral airflow, pharyngeal and oral pressure changes were recorded using a mouth mask connected to pressure/flow transducers. A pressure tube was inserted behind the lips at the right edge of the mouth. For the detection of *overlapped* articulations, the methodology is quite complex but basically centres on the oral pressure output from front-to-back vowel transition such as [ipku]. As the tongue body moves backwards for this vowel transition, labial and velar closures occur simultaneously for *pk* and seal the oral cavity at both ends at the lips and at the soft palate. The tongue body backing motion across the soft palate expands this sealed oral cavity which leads to negative pressure. Therefore the resulting pressure rarefaction is indicative of a highly overlapped labial to velar sequence. For the detection of reduced articulations, oral pressure output can distinguish a no-change in pressure state or a change in pressure state. If the labial closure for *pk* is not achieved then no pressure builds up behind the lips.

Korean is a language which, like English, permits optional assimilation of coronals to a following velar or bilabial. For both languages this is described as an optional phenomenon motivated by casual/fast speech. Unlike English, labial to velar assimilations are also permissible in Korean.

Two main experiments were carried out, the production and perception of Korean *pk* sequences, production and perception of English *pk* sequences. V_1p#kV_2 sequences embedded in meaningful sentences were produced by 14 Korean speakers and 8 English speakers where V_1 and V_2 represent front and back vowels respectively. Velar –velar control sequences were also recorded. The Korean speakers read the sentences 3 times in a formal speech style and 3 times in a casual speech style. The English speakers read them 3 times in a formal style only.

Four different patterns of oral pressure were yielded from the recordings (too detailed to describe here) which highlight the different articulatory strategies employed by speakers to coordinate the labial to velar clusters. The different patterns of oral pressure were taken to correspond to (i) slight/no overlap (ii) marked overlap (iii) reduction (iv) extreme overlap. Some of these strategies turned out to be somewhat speaker-specific.

Results for production of Korean and English *pk* are shown in Tables 1.5 and 1.6 below, note that for English only formal speech tokens were produced. Results for the production of Korean *pk* suggest that reduction of the bilabial is by far the preferred strategy for casual speech and also overall.

Table 15 Korean results for /pk/ *overlap and reduction types across formal and casual speech (adapted from Jun, 1996: 387)*

KOREAN:

	p | k slight/no overlap	**p^k** marked overlap	**<p>k** reduction	**k^p^k** extreme overlap	**k^p** (error)
formal	23 (14%)	77 (47%)	51 (31%)	12 (7%)	0
casual	29 (17%)	31 (18%)	105 (63%)	1 (0.6%)	1 (0.6%)
total	52 (16%)	108 (33%)	156 (47%)	13 (4%)	1 (0.3%)

Table 1.6 English results for /pk/ *overlap and reduction types, formal speech, 8 speakers (adapted from Jun, 1996: 396)*

ENGLISH:

subject	**p | k** slight/no overlap	**p^k** marked overlap	**<p>k** reduction	**k^p^k** extreme overlap
1	1	14	0	0
2	8	7	0	0
1	1	14	0	0
2	4	10	0	1
3	9	6	0	0
4	10	5	0	0
5	1	14	0	0
6	4	11	0	0
total	38	81	0	1

Jun also found inter-speaker variability of pressure patterns for Korean. 5 of the 14 speakers employed the reduced labial strategy for *all repetitions* in casual speech whereas in the same condition 4 speakers produced 'marked overlap' (*p^k*) tokens for more than half of all repetitions. It seems the reduced labial strategy is dominant but not for all speakers.

Looking at the English results, in Table 1.6, only three oral air pressure patterns were observed. No *<p>k* outputs were found, not surprisingly, the labial closure was never reduced in English for *pk*. Most patterns reveal either slight/no overlap or marked overlap. The latter of these patterns is twice as frequent as the former, 81 tokens versus 38. Jun comments:

> Consequently, English *pk* clusters are not very different from Korean *pk* clusters in gestural overlap, since both clusters are usually highly overlapped (67% of English test tokens; 70% of Korean test tokens with unreduced /p/). In contrast, English *pk* clusters are completely different from Korean *pk* clusters with respect to reduction of *p,* in that the *p* never reduces in English, but often does in Korean (47% of all Korean test tokens). p.395

5 of the eight 8 English speakers tended to opt for a 'marked overlap' strategy while three of them opted for a more even distribution of 'marked overlap' and 'slight overlap' and thus were more variable.

For the perceptual part of this experiment 6 Korean-speaking listeners heard extracted experimental sequences from the data collected earlier, selected on the basis of air pressure measurements. The listeners were asked to decide whether they heard V_1pkV_2 or V_1kkV_2. Jun found that:

> ...in Korean labial-velar consonant clusters, labial reduction obscures the perception of the labial, but a marked overlap of the consonant cluster does not in general obscure the perception of the labial unless labial reduction, partial or complete, is involved. *In other words, it is gestural reduction, not gestural overlap, that plays a decisive role in the perceptual assimilation of Korean labials.* (p.394: italics added)

While possible problems with the methodology have been acknowledged by the author, this study has usefully presented counter-evidence to the claim made by Articulatory Phonology that the timing between or 'phasing' of individual gestures alone is sufficient to result in a perceived assimilation. Care must be taken, however, in generalising these findings to other sequences, such as alveolar to velar sequences which involve different parts of the *same* articulatory sub-system, and in generalising to sequences which are tauto-syllabic.

1.4 LANGUAGE/DIALECT SPECIFIC ASSIMILATION AND PHONETIC RULES

The question of variability in the production of alveolar to velar sequences should be seen within the broader context of place assimilation as a language/dialect specific phenomenon. Provisional confirmation that alveolar to velar assimilation in Russian is nowhere near as extensive as it is in other languages has been provided in an EPG study by Barry (1988). Furthermore, Farnetani and Búsa (1994) found that in Italian, assimilation in /nk/ clusters is always categorical. They analysed EPG patterns of 7 repetitions of test items from each of 3 speakers. On the basis of an auditory study of Durham English, Kerswill (1987) notes either an absence or a near absence of place assimilation in contexts where it might reliably be predicted to occur in other varieties of English.

Language-specific effects are also found in 'lower level' coarticulatory phenomenon. Solé and Ohala (1991) found that anticipatory vowel nasalisation in Spanish is phonetic while in American English it is phonological. As a function of speech rate they found a quite different distribution of the timing patterns of nasalisation in the two languages. In Spanish the extent of nasalisation on the vowel which precedes a nasal stop remained constant while for American English it was always proportional to varying vowel duration. They concluded from this that as vowel duration increases in American English the spread of nasalisation is controlled by the speaker, that is, is phonological but in Spanish the fact that the extent of nasalisation is always constant means that this equals the time it takes for the velum to lower for a nasalisation target. Cohn (1990) compared the nasal air-flow contour of nasalised vowels in English, French (where nasal vowels are phonological) and Sundanese (where vowel nasality is described as due to a feature-spreading rule). In the Sundanese vowel, the air-flow patterns have plateau-type shapes very similar to the French patterns. But the English shapes are smooth transitions from closed velum to open velum throughout the duration of the vowel. Thus nasalisation is categorical in French and Sundanese but gradient in English. Öhman's classic study of VCV coarticulation in Russian (1966) showed that the tendency in English for the tongue to start assuming the position for V_2 during and even before the consonantal closure was not found for Russian. This was attributed to the fact that Russian contrasts palatalised and nonpalatalised consonants and so the tongue has less freedom to anticipate upcoming tongue targets.

Evidence that place assimilation can be thought of as language/dialect-specific has important ramifications for internal models of speech production. Nolan (1992: 278) noted that if it does become established as a phenomenon over which speakers have control, 'this will provide evidence that a greater amount of phonetic detail is specified in the speaker's phonetic representation or phonetic plan than is often assumed.' Assuming that high-level phonological information is minimal and is not characterised by more-or-less values, detail must be handled at lower phonetic levels. Furthermore, as long as individual languages/dialects constrain the occurrence of connected speech processes such as assimilation these details must be learned and incorporated into the grammar. An important consequence of this is that gradient assimilatory forms cannot be attributed to the properties of the vocal mechanism alone as some have argued (Zsiga, 1995, Holst and Nolan, 1995).

At this point it is necessary to outline what might be considered a standard psycholinguistic model of speech production within which various processing levels can be identified. Levelt's influential model of speech production (1989) is typical of models which separate out phonological and phonetic processing levels. It is generative in principle and we can assume that within this model coarticulation, assimilation and connected speech processes are all stages which have a bearing on the relationship between competence and performance. Levelt's model encompasses all stages of production including syntactic processing but what concerns us here is specifically those processes from the level of lexical access down. First a level of *phonological encoding* is proposed whose function it is to retrieve or build an articulatory plan for the utterance: 'The major source of information to be accessed by the Phonological Encoder is *lexical form,* the lexicon's information about an item's internal composition…an item in the lexicon contains information about its morphology and its phonology.' (p.12) At this level, many phonological procedures will be applied which either modify or further specify information that has been retrieved. Phonological encoding gives rise to the next stage called the *phonetic* or *articulatory plan.* This is an internal representation or 'programme' for articulation, not articulation itself: 'internal speech is the phonetic plan as far as it is attended to and interpreted by the speaker.' Following this comes *articulation* 'overt speech' which is the execution of the phonetic plan by the physical systems of speech. Levelt proposes that the generation of internal speech is generally ahead of speech execution and this is facilitated by the *Articulatory Buffer.* The articulator (the module responsible for articulation) can retrieve chunks of internal speech located in this buffer and implement them for actual speech.

Fourakis and Port (1986) have advanced an argument in favour of enhancing a particular role that the phonetic level has in the derivation from abstract phonological representation to actual speech output. They looked at the phenomenon of stop epenthesis in English, the insertion of a stop between a sonorant and a fricative consonant in syllable-final sonorant-fricative clusters. Traditionally, stop epenthesis as in, for example, some realisations of *prince* or *else* has been explained either as following from universal constraints on the speech production mechanism or as the result of language or dialect-specific phonological insertion rules stated in the grammar. Therefore these two theories locate epenthesis at the abstract planning stage and the concrete output stage of the speech chain respectively in a very similar way in which place assimilation is located by competing theories. Ohala (1974) has proposed that in the case of nasal-fricative clusters as in *tense* a stop will occur quite naturally as a result of mistiming from carelessness or a lack of fine motor control. He claims that if the velum is closed before the oral stop for the nasal is released then a configuration will come about similar to that for a homorganic stop. Carelessness also comes up in Ohala's hypothesis that assimilation has its roots in the misperception of VC_1C_2V sequences as having only one consonant place of articulation.

On the basis of their acoustic measurements of data collected from a group of South African English speakers and a group of American English speakers, Fourakis and Port conclude that the source of stop epenthesis is neither phonological nor phonetic in nature. Their subjects were recorded saying the test items such as *tense, dense, false* in the carrier phrase 'Bob said_____today' as well as control items such as *tents, dents and faults*. The purpose of the controls was so that if epenthetic stops did occur, they could be compared with the lexically underlying forms. They had two main findings. Firstly, the South African English speakers always maintained a clear phonetic surface contrast between pairs like *tense* and *tents*. They invariably produced sonorant-fricative clusters with no intervening stop and sonorant-stop-fricative clusters with a full stop. The American English speakers, however, did epenthesise the sonorant-fricative clusters. This finding suggests that a universal mechanical basis for this phenomenon is wrong. The second main finding is that the inserted stops produced by the Americans are shorter in duration compared to their underlying stops ('incomplete neutralisation') and are thus articulatorily distinct. So it seems that speakers learn to insert the epenthesis 'in a completely consistent fashion' (p.220) and yet the duration evidence points to the fact that if speakers were operating a segmental insertion rule the inserted item would have to be segment sized, which, it turns out, it is not.

Rather than invoking and modifying different models of speech production to account for these results, Fourakis and Port propose a special type of rule called a *phase rule*. This is essentially a language-specific (i.e. learned) low-level rule which governs articulatory transitions between neighbouring segments and in this sense would seem to be fruitfully applied to place assimilation. It governs local articulatory gestures which includes the timing of these gestures relative to each other and the result is graded parameters of articulation. This then accounts for cases of incomplete neutralisation as their data indicates. But what makes this rule not entirely phonetic is that it has access to phonological information such that alterations in articulatory timing occurs only in a specified environment, in this case syllable-final sonorant-fricative clusters. They also say that this phase rule strongly resembles post-lexical rules. In the case of alveolar assimilation, target undershoot forms of alveolar stops suggest that low-level inertial effects are at work and yet this is permissible only in certain phonetic contexts and in certain languages/dialects.

The location then, of these phase rules is an intermediate structure 'between the language (that is, the phonological representation) and actual gestures.' (p.199) Rather than *all* phonetic implementation rules being directly controlled by speakers (that is, learnt by speakers as language/dialect-specific rules) they consider it 'plausible that the speech production system is arranged so that a fairly small number of parameters may be directly controlled by linguistic factors' within this intermediate level. These are governed by the phase rules which alters timing detail in the appropriate linguistic context.

1.5 PERCEPTUAL AND ACOUSTIC STUDIES OF STOP CONSONANT SEQUENCES

Byrd (1992) has tested the perceptual effects of gestural overlap/coproduction on C_1 of $VC_1\#C_2V$ sequences using an articulatory speech synthesizer. The speech synthesizer is a computational model based on the theoretical ideas of gestural phonology. Input to it is the gestural score in conjunction with the task dynamic model which calculates the patterns of articulator motion. The general aim was to investigate the relationship between assimilation and coproduction/ coarticulation in a gestural model. That is, whether listeners perceive consonant to consonant sequences as assimilated if the amount of gestural overlap between them is artificially manipulated. It is already known that the coproduction of two consonants across a word boundary in $VC_1\#C_2$ sequences has formant transitions in the preceding vowel which are characteristic of the place of constriction for C_2, since in highly coproduced utterances, the articulators are already moving to form the target for C_2 at the onset of the vowel. Byrd's hypothesis, then, is that listeners will use the acoustic information resulting from coproduction in order to identify C_1. By contrast, Jun

(1996) has proposed that overlap alone is not enough to produce the percept of assimilation.

Gestural overlap only was used by Byrd to model coarticulation causing the percept of assimilation - gestural magnitude was not included as a variable. The test stimuli phrases were synthesized with consonant clusters /b#d/; /d#b/; /b#b/; /d#d/ where the final two are controls. Byrd considers that the underlying gesture will always remain in the output even in the context of substantial assimilatory effects and argues against the proposal that in some cases the alveolar gesture in alveolar to velar sequences disappears altogether. Gestures are organized in time chiefly by the relation between 'gestural activation periods'. These periods are *phased* and so gestures are coordinated serially in different ways according to the context.

The experimental stimuli consisted of 11 grades of overlap for each of the four phrases. The four phrases were:

/bæb#bæn/ bilabial # bilabial (control, no assimilation)
/bæb#dæn/ bilabial # alveolar
/bæd#bæn/ alveolar # bilabial
/bæd#dæn/ alveolar # alveolar (control, no assimilation)

The acoustic and articulatory correlates of overlap were defined as the movement of formant transitions and length of closure interval. The closure interval of the stimuli was silent and, importantly, it did not include release bursts for either C_1 or C_2. A very few tokens (five) were not overlapped sufficiently to prevent a release period for C_1. Interestingly she found that when the amount of overlap was gradually increased from a small amount (where C_1 was unambiguously released) the alveolar release of /alveolar # bilabial/ sequence was more robust than the bilabial release of the /bilabial # alveolar/ sequence. That is, when the percentage gestural overlap was raised to 56% for the sequences, the /d#b/ sequence still yielded a perceptual release whereas for /b#d/ release did not occur above an overlap percentage threshold of 34%: 'The fact that C_1 is released at different phasings for different places of articulation reflects the fact that different gestures entail different speed of release, or different effective stiffnesses in movement away from their targets.' (p. 7) The tongue tip articulator moves more rapidly than the lip articulator. When there is so little overlap that C_1 is released, it is expected that no C_2 acoustic effects on the preceding vowel are available to the listener making the perception of assimilation of C_1 unlikely.

The results seem to support Byrd's hypothesis that listeners do attend to the acoustic effects of gestural overlap. Listeners utilise formant transition information, relative place of articulation and closure interval to judge the place of articulation of the consonant closure and whether assimilation had occurred, rather than utilizing the presence or absence of release. She infers from this that the absence of a release is not automatically indicative of coproduction or assimilation and she draws a parallel with this fact and previous work which has established that perception of both consonants in a VCCV sequence is typically accurate even though no medial release phase is present, a commonplace effect in natural speech.

The percepts of the two non-homorganic clusters /b#d/ and /d#b/ both changed as a result of increasing gestural overlap and this was a highly statistically significant effect. For /b#d/, judgements of assimilation first start to rise above 10% of all responses (i.e. assimilated, 'ad' or unassimilated, 'ab') when gestural overlap is at 78%. However, judgements of assimilation do not rise above 20% until there is as much as 100% overlap and for this sequence, perception of C_1 as an alveolar is never more frequent than its perception as a bilabial (unassimilated). In contrast to this, the /d#b/ sequence results show that the frequency of assimilation judgements begins to rise *rapidly* at 63% gestural overlap and when phasing was at maximal overlap this /d#b/ cluster was considered bilabial 88.5% of the time.

The obvious differences in assimilation responses to these clusters clearly says something about the special status of coronals i.e. their particular vulnerability to the effects of adjacent speech units. Her rationale in acoustic terms is that the acoustic consequences of constrictions overlapped by bilabials and alveolars 'differ in the response of the vocal tract to these places of constriction due in part to the greater intrinsic stiffness or velocity of the tongue tip gesture. The asymmetry obtained in this perceptual experiment was found to conform to assimilation processes in natural connected speech.' (p.21) Additionally, she says that the magnitude of formant transitions for alveolar and bilabial closures differ - generally alveolars show smaller formant transitions than bilabials (she cites Öhman, 1967; Stevens et al, 1966). Velar and bilabial movement is slower and involves a larger mass of tissue than an alveolar movement and thus might be more difficult to hide by an upcoming consonant. It is possible that alveolars are often eclipsed partly because this can be done with relative ease.

Byrd argues that gestural overlap is indeed the correct way to model assimilation and that the gestural model she uses offers the 'additional advantage of representational simplicity because assimilation is modelled as a direct consequence of the input representation rather than any rule, action, or condition altering the input representation.' (p.18). This position also gets around the unsatisfactory shifting of the problem of the gradient nature of place assimilation to different levels of the grammar (as witnessed in recent feature geometry and autosegmental accounts). Whereas a coproduction model merely yields mechanical gradation of articulations only, this model is able to reflect language-specific types of place assimilation because canonical phasing rules are language-specific. Indeed it is claimed that the model can make some predictions about the nature of specific language differences in connected speech processes. For instance, a language that has a small degree of canonical overlap for a certain cluster of consonants will demonstrate fewer assimilations than a language which for the same cluster demonstrates a greater canonical overlap.

Another study of the relationship between perception and assimilation has been carried out by Hura, Lindbom and Diehl (1992). Their argument is that the role of perception has been overlooked in the search for explanations of assimilation. The aim of their experiment is to test Kohler's hypothesis (1990) that fricatives do not assimilate because members of the fricative class are perceptually distinct. By contrast, unreleased oral stops and nasal stops do assimilate because they are relatively more confusable between themselves and so an assimilation would not be particularly salient. Kohler observes that assimilation of alveolars occurs syllable-finally in German: a<u>nb</u>ringen [mb]; mi<u>tb</u>ringen [pb], mi<u>tg</u>ehen [kg] but not syllable initially. Kohler comments:

> ...the initial position in a word has high signalling value for a listener and thus demands greater articulatory precision from a speaker than the final position...fricatives are not assimilated under any conditions, because they are acoustically and auditorily far more distinct than nasal and unreleased stops with regard to place cues so that their articulatory reduction would be too salient, and is, therefore, not tolerated. (p.88)

Hura et al investigated the possibility that English fricatives are in fact more accurately identified in the context of a following stop consonant than either nasals or unreleased oral stops. Perceptual stimuli was created from recordings of 8 speakers of American English. 22 speakers of American English listened to the stimuli and were asked to make judgements. The results showed that nasals had the highest error rate (6.9%), followed by unreleased stops (5.6%) and fricatives (3%). This result led Hura et al to propose that perceptual factors play a role in the on-line production of assimilations. A segment's distinctiveness is crucial to its phonological behaviour.

An additional perspective on assimilation has been provided by acoustic and perceptual studies of place cues. Malécot (1958) found that the released and unreleased consonants of heterorganic clusters contain different types of information. He spliced the syllable /ep/, minus the stop release, to the release of the stop in the syllable /ek/ to yield /epk/ and showed that listeners overwhelmingly identified this as only a single consonant with the place of articulation of C_2 (/ek/). He concludes that:

> Voiceless p t k releases and voiced b d g releases contain sufficient cues for conveying both place and manner of articulation of American English plosives in final position. These cues are powerful enough, in most instances, to override all other place and manner cues present in the vowel-plus-closing transitions segment of those plosives. (p.380)

In the case of final unreleased stops then, only stop *onset* cues are present. Not only are these less utilisable perceptually (at least in comparison to final released stops where transitional *and* release information is present) but also give rise to ambiguous percepts in terms of place. In VC environments, stop onset cues are often misidentified (Householder, 1956). It would not be unreasonable to suggest, then, that stops in post-vocalic position are more likely to assimilate than when in pre-vocalic position because the cues are weak here anyway.

Importantly, nasal stop cues have been found to be even weaker still than non-nasal place cues. Nasals are known to be confused amongst themselves (House, 1957). The nasal/non-nasal opposition is highly distinct however. For instance, in the case of the phrase /braun#kəʊt/ *brown coat* realised as /brauŋ# kəʊt/, the most likely perceptual outcome given this hierarchy, is that the /n/ will be judged nasal but will take its place cue from the stop release /k/. It is possible that this is the perceptual default setting for realisational differences in affected segments although it is unclear how partially assimilated forms function. Conversely, in pre-vocalic position, place cues for stops are very robust. It has been argued that most of the time the burst alone provides a sufficient place cue (Winitz, Scheib and Reeds, 1972) since the spectrum of the turbulence is generated by the oral tract configuration forward of the constriction.

Byrd (1992) has a very different perspective on the locus of important acoustic information, as described in section 1.5 above. She contends on the evidence obtained by Walley and Carrell (1983), amongst others, that CV formant transitions are:

> at least as important a cue as the consonantal release bursts. They [Walley and Carrell] presented synthesized stimuli with conflicting burst spectra and formant transitions and found that the identification of the consonant corresponded much more reliably to the formant transitions than to the bursts. On the basis of this evidence, we can assert that the type of assimilation favoring C_2 in VC_1C_2V sequences as described by Ohala and others can also be a function of the salience of the onset formant transitions of the vowel following closure, rather than a sole consequence of place information found in the release burst. (p.13).

1.6 RESEARCH QUESTIONS

This survey has shown that there is a considerable body of literature on place assimilation and related phenomena. Some interesting issues have emerged that were considered in need of further investigation. One of the biggest shortcomings of previous work on alveolar to velar place assimilation, acoustic and EPG work, is the small number of speakers involved and the small number of repetitions elicited from each. This situation, combined with the lack of focus on the assimilatory behaviour of individual speakers, has given rise to the assumption that all speakers assimilate this sequence gradually, that is, produce a continuum of assimilatory forms including partial assimilations. The experiment reported in the following chapters is a larger scale investigation of alveolar to velar assimilation with the question of distribution of residual alveolars at the forefront. Another motivation for the study was that little work has been carried out on the articulatory details of alveolar nasal stops. The investigation has two parts. Firstly an EPG experiment is described involving 10 speakers, who each produced 10 repetitions of experimental sequences. Following this a smaller scale, combined EPG/EMA experiment was carried out which was designed to follow-up some findings of the first experiment. The research questions which underpin this investigation are listed below:

Question 1

How do speakers produce the sequence /n#k/?

Specifically in terms of:

- effect of speech rate on spatial/timing description of the experimental sequence.
- amount of inter-speaker variation.
- amount of intra-speaker variation (individual speakers' assimilatory variation across repetitions).
- possibility of phonetic identity of underlying /n#k/ sequences and all-velar /ŋ#k/ control sequences.

Question 2

At what level in the generation and execution of an utterance does assimilation occur?

Question 3

How do the articulatory details of assimilation match the existing models?

Question 1 is intended to shed light on the types of assimilation subjects produce and their frequency. In the data analysis, particular attention will be paid to individual speaker trends. In cases of apparent complete alveolar assimilation, as defined by EPG patterns, traces of the alveolar gesture will be looked for. If none are found and underlying /n#k/ sequences are indistinguishable from control /ŋ#k/ sequences, then these cases of /n#k/ assimilation will be interpreted as involving a discrete switch in phonemic identity. However, the process responsible for this discrete change (put bluntly, from /n/ to /ŋ/) will need careful consideration. Holst and Nolan (1995) assert that /s#ʃ/ assimilation yields a continuum of forms including partial loss and complete loss of the /s/ gesture (the latter being their 'type D' forms, see section 1.3.5). Holst and Nolan, however, consider these type D forms to be motivated by an explicit cognitive categorical rule. Gestural reduction, they propose, increases until a threshold is reached which somehow activates this cognitive rule. The assumption we are forced to make, tested in the experiment reported in this dissertation, is that all speakers assimilate in this way, producing a continuum of forms and that this cognitive rule can only be activated via this threshold. Previous EPG studies of assimilation (Hardcastle and Roach, 1979; Wright and Kerswill, 1989; Barry, 1985) that had also noted this continuum had attributed the complete assimilatory forms to maximal reduction only, located at the end of the reduction continuum, thus viewing assimilation as a wholly phonetic process. An important aim of the study reported in this dissertation is to find out if this is the case or whether categorical changes for a particular speaker can occur without evidence of partial assimilatory forms.

Answers to Question 1 can next be related to the question of *where* assimilation occurs in the generation and execution of an utterance (Question 2). Finally, the various models of speech production outlined in section 1.2 of Chapter One will be evaluated in terms of their capacity to account for the results of this study.

CHAPTER TWO

Methodology

This chapter is in two parts. Part 1 (section 2.1) describes the methodology for the main EPG experiment and Part 2 (section 2.2) describes the methodology for the follow-up EPG/EMA experiment.

2.1 EPG METHODOLOGY

This section describes the methodology for the EPG-only study in the following order:

1. The test material and subjects
2. Elicitation procedure
3. Instrumentation and recording conditions
4. Method of EPG and acoustic data annotation
5. Methods of data analysis, timing and spatial measures

2.1.1 Test materials

Test materials were devised so that a word boundary occurred between the two elements of the experimental sequence. The sequences were:

(i) alveolar nasal to velar plosive, **/n#k/**

(ii) velar nasal to velar plosive, **/ŋ#k/**

The sequence /n#t/ was not used as a further control because the focus of attention is on assimilatory changes in /n/ and not more subtle changes in place of articulation that can occur in non-assimilated /n#k/ sequences compared to /n#t/ sequences.
Sequence (i) above captures the potential site of alveolar assimilation and (ii) is a control sequence which provides a means of comparing /n#k/ assimilatory forms with the lexical velar form /ŋ/.

Meaningful sentences were devised to carry these experimental sequences. A further 8 ‘filler’ sentences of no experimental interest were added to the 2 experimental sentences.

These fillers were added as an attempt to distract the subjects from the presence of near-minimal pairs in the experimental sentences. Table 2.1 below shows all test material:

Table 2.1 complete list of experimental and filler sentences used in the EPG experiment

	experimental sentences
1	It's hard to believe the ban cuts no ice
2	I've heard the bang comes as a big surprise

	'filler' sentences
1	Just imagine the noise coming from that place
2	He told them I'd jump at the chance
3	I'm sure they'll accept it this time
4	We need a fan, a desk and a new chair
5	Some parts of the city are really noisy
6	Do you think newsreaders use too much slang?
7	I think he still smokes
8	They've even stopped them doing overtime

Care was taken in the construction of the word carrying the word final /n/ or /ŋ/ to avoid a word-initial lingual stop or lingual fricative articulation. Words beginning with the non-lingual plosive /b/ were used so that there would be minimal carry-over coarticulatory influence on the production of the experimental alveolar or velar stop. The vocalic environment was kept as consistent as possible. Bordering vowels were /a/ & /ʌ/. Low vowels were used to flank the consonants to avoid vocalic tongue-palate contact (characteristic of high front vowels for instance). Also it was predicted that a relatively front velar place of articulation would be achieved after /a/, making identification of velar closure easier. Voiceless plosives followed the nasal in all experimental sequences so that the offset of voicing would facilitate identification of the offset of the nasal stop and beginning of the voiceless phase of oral closure.

Adobe Photoshop Elements Clip Image is too big to be exported

2.1.2 Subjects

10 Subjects with EPG palates were recruited from the Department of Speech and Language Sciences at Queen Margaret University College, Edinburgh. No subjects reported any speech or language pathology or hearing impairment. Most subjects were experienced wearers of EPG palates and had participated in previous EPG experiments carried out in the Department. The point must be made here that since all subjects were recruited from an academic speech department, they could not have been naïve to the fact that their production of consonants was under investigation. 8 of the subjects were female and 2

were male and their ages ranged from 24 - 40yrs. The subject group overall represented a range of accents. Four spoke Standard Southern British English (subjects D, H, I and J); one subject was originally from the North East of England (subject F); one was originally from Eastern Australia (subject G); one spoke with a Northern Irish accent (subject C) and the remaining three subjects were Scottish. Two of these three were from the West of Scotland, Glasgow (subjects A and B) and the other was from Fife, near Edinburgh (subject E). There was no attempt to control for accent whereby all subjects would have the same accent. This was because there is no evidence in the literature that the frequency and type of assimilation that /n/ can undergo in an assimilating context is accent-specific. No speaker had the type of accent (e.g. Midlands, U.K.) which features a velar stop reflex following word-final /ŋ/ e.g. [baŋg kʊts].

2.1.3 Elicitation Procedure

Elicitation of the data fell into two parts. The goal of the first part was to elicit slow and careful speech and the goal of the second was to elicit fast and casual speech, although genuinely casual speech was not realistically expected from laboratory conditions.

In the slow and careful speaking condition 10 repetitions of all 10 sentences shown in Table 2.1 above were required from each subject. This meant that in this part of the recording, subjects each produced 100 slow/careful speech sentences. Subjects were presented with a randomised list of single sentences numbered 1 to 100. Before recording started, subjects were instructed to practice reading two or three of the sentences slowly and clearly to check that speech rate was not excessively slow. No instruction was given with respect to prosodic structure of the material.

For the second part of the experiment subjects were required to produce 10 repetitions of the same 10 sentences but in a fast/casual mode of speech. This time however, the list they were given to read from did not comprise of numbered single sentences. The stimuli were organised in the following way as an attempt to elicit relatively naturalistic fast/casual speech and to some extent control the speed of subjects' speaking rate. The original 100 sentences were randomised and ordered in groups of 3 with filler sentences purposely distributed to avoid clustering of experimental sentences. This meant that each group of 3 contained one experimental sentence. Because 100 is not exactly divisible by 3 two filler sentences were repeated once again to make up a further group of 3, bringing the number of groups up to 34. Subjects were asked to read out the sentences in each group one after the other avoiding pauses in between. A time limit of 5 seconds was imposed on the production of each group of 3 sentences. Subjects were asked to practice reading two or three groups while being timed. An example of the materials subjects were given is shown below in Figure 2.1:

1

Just imagine the noise coming from that place.
It's hard to believe the ban cuts no ice.
Do you think newsreaders use too much slang?

2

They've even stopped them doing overtime.
We need a fan, a desk and a new chair.
I've heard the bang comes as a big surprise.

Figure 2.1 sample fast speech stimuli for EPG experiment – stimuli was presented to subjects in groups of three as shown

2.1.4 Instrumentation and recording conditions

The Reading electropalatography system (EPG3) was used to acquire the data. This records an acoustic signal along with details of the timing and location of tongue contact with the hard palate. The artificial palate worn by subjects and the EPG 3 system will be described below. Further details of the system can be found in Hardcastle, Gibbon and Jones (1991).

2.1.4.1 The artificial EPG palate

Artificial custom-made palates were made from a full impression of subjects' upper teeth and hard palate. All palates were made by Broughton and Tyrell of Newbury (UK). 62 silver electrodes, 1.4mm in diameter, are embedded into the palate. The palate itself is made of acrylic and is approximately 0.8mm thick. Each electrode is joined to a piece of enamelled copper wire (41swg) and channeled to the posterior corners of the palate. They are covered in a soft heat shrinking tubing and come out from the mouth via the buccal surfaces of the maxillary dentition (see Figure 2.2 below).

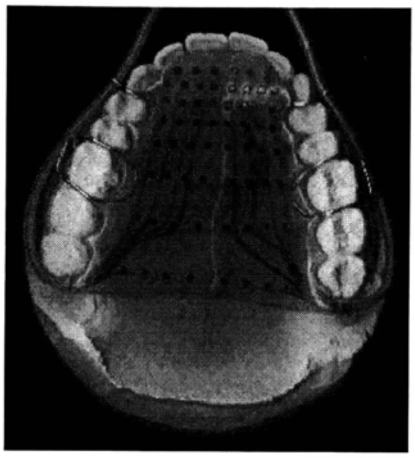


Figure 2.2 EPG palate and dental impression (taken from the University of Reading EPG pages: http://midwich.reading.ac.uk/research/speechlab/

The electrodes are arranged into eight rows with eight electrodes in each row apart from the first row which has six electrodes. Placement of the electrodes corresponds to predetermined anatomical areas. These areas are phonetically relevant in terms of place of articulation, for example the junction between the hard and soft palate. The alveolar region has a relatively large concentration of electrodes since several speech sounds are alveolar in English. The first 3 rows of the palate correspond to the alveolar region, rows 4 and 5 correspond to the palatal region and rows 6, 7 and 8 correspond to the velar region. (see Figure 2.3 which shows a schematic representation of the palate).

0 0 0 0 0 0	
0 0 0 0 0 0 0 0	ALVEOLAR REGION
0 0 0 0 0 0 0 0	
0 0 0 0 0 0 0 0	PALATAL REGION
0 0 0 0 0 0 0 0	
0 0 0 0 0 0 0 0	
0 0 0 0 0 0 0 0	VELAR REGION
0 0 0 0 0 0 0 0	

Figure 2.3 Division of EPG palate into 3 different articulatory regions (solid line) mid sagittal area indicated by dashed line

There are some concerns regarding the acquisition of normal sounding speech using EPG. Hardcastle (1972), however, suggested that the presence of the palate during speech does not cause interference with normal articulation and that patterns of tongue-palate contact produced with electropalatography were not significantly different from those produced with direct palatography. Electrodes on the artificial palate cover the hard palate only and so there is an issue concerning the representation of contact for velar closure. The representation of contact for 'front' velars such as those produced before a high-front vowel, however, is not a serious problem.

2.1.4.2 EPG3

The processor used to run EPG 3 was a DCS 486 PC. The subject during recording wears the palate which is plugged into a multiplexer unit worn around the neck. An electrode attached to the multiplexer is held by the subject which creates a circuit. If this electrode is not held by the subject then no EPG patterns will be recorded. The electrode generates a low-frequency (15Hz) sinusoidal signal with a magnitude of 300mV RMS. If a signal is present as electronic circuits scan the 62 silver electrodes then tongue contact with the palate is registered. This contact information is then transmitted to the computer for storage and display. Figure 2.4 shows the EPG 3 system.

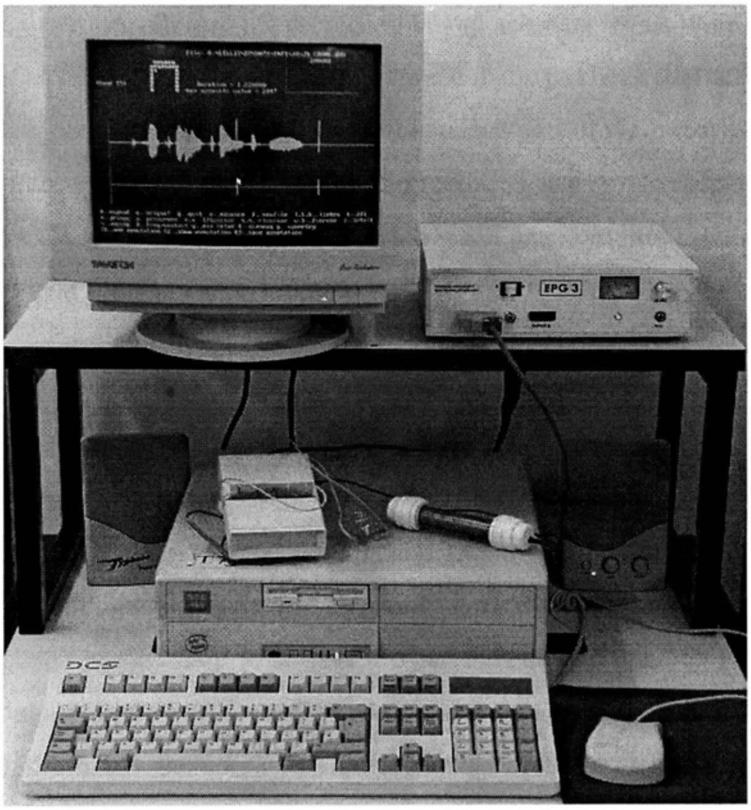


Figure 2.4 EPG3 system showing multiplexer, artificial palate and hand-held electrode

EPG data is sampled at a rate of 100 frames per second and the acoustic signal at 10Hz. The waveform and EPG data is synchronised. A single microphone (audio-technica AT

803B) was used to simultaneously record the EPG acoustic signal and the acoustic signal for the digital audio (DAT) recording. The DAT recorder was a Sony DTC 60ES.

2.1.4.3 Data collection

Recordings were made in a single session usually lasting approximately 30 minutes. Subjects were required to wear their palates for at least an hour prior to recording to make sure they acclimatised to the feel of it in the mouth. The subject was seated in a sound damped studio in the laboratory of the Department of Speech and Language Sciences. The only piece of equipment left inside the sound damped studio with the subject was the multiplexer worn around the neck and the microphone. The processor and the DAT recorder were in a control room to so that any noise generated by this equipment was excluded from the recording. A small window divides the studio and the control room through which the subject can see the researcher and instruction signals throughout the recording session. Before the recording began acoustic levels were set for both the EPG and the DAT recorder. Test recordings were made to check that all electrodes were registering contact. For the slow/ careful speech part of the experiment, recorded first, each sentence was individually cued and recorded as a single speech file. For the second part of the experiment, fast/ casual speech, each group of three sentences was recorded as a single speech file.

2.1.5 Method of data annotation and analysis

Prior to formal annotation of the data it was checked through using the screen display to get an overview of some basic tendencies. In particular, frequency of alveolar to velar assimilation for each subject was checked. An example of a screen display for the careful speech sentence 'It's hard to believe the ban cuts no ice' is given in Figure 2.5. On the screen display there are two independently-moving cursors. Wherever the left-hand cursor is placed along the waveform the palate display in the top left-hand corner will show the details of the tongue-palate contact for that moment in time.

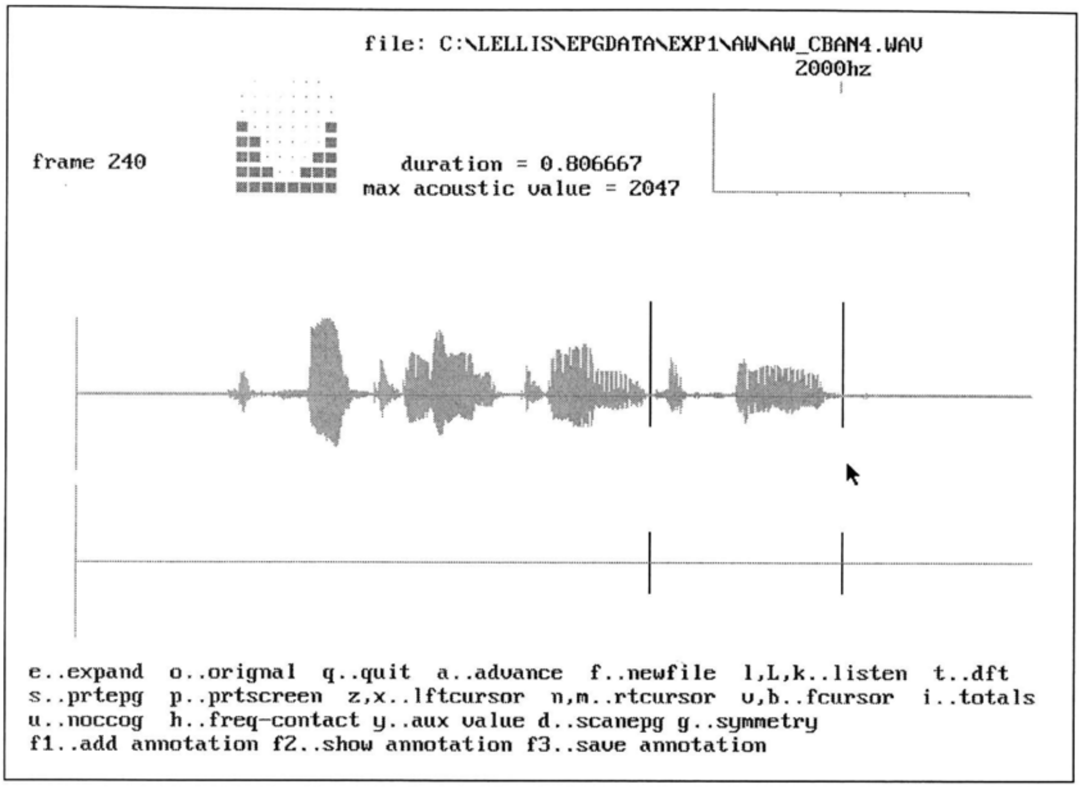


Figure 2.5 Screen display of 'It's hard to believe the ban cuts no ice' *careful speech. Frame display at top left shows velar closure prior to release for /k/.*

Because the two cursors move independently, the data can be segmented and a print out of the selected area can be obtained. Figure 2.6 shows an example of printed EPG data. The tongue-palate contacts show alveolar stop closure followed by velar stop closure from a careful production of …*ban cuts*… The data is read from left to right with each EPG frame 10ms apart.

2.1.5.1 Annotation procedure

Electropalatographic measurements were taken from the EPG trace and acoustic measurements were taken from the waveform. No instrumentation was used to measure the raising of the velum and so the transition from nasal to oral airflow during the experimental sequence was not annotated. For all experimental sequences, annotation points were made between the onset of the vowel before the word boundary (/a/ in all cases) up to onset of the vowel after the word boundary (/ʌ/ in all cases), i.e. /ban kʌts/ or /baŋ kʌmz/. A list of annotation points, their type and description, is shown in Table 2.2

Table 2.2 list of annotation points: acoustic and EPG

label	*type*	*description*
v1o	acoustic	Onset of periodicity for the vowel /a/
cl1	EPG	Onset of mid sagittal contact in first 3 rows for stop closure /n/
cl2	EPG	Onset of complete or maximum constriction for stop closure /k/
re1	EPG	Earliest appearance of loss of mid sagittal contact for stop closure /n/
env	acoustic	Onset of voiceless phase of oral closure for /k/
re2	EPG	earliest appearance of loss of maximum or complete constriction for /k/
v2o	acoustic	Onset of periodicity for the vowel /ʌ/

All annotations were made in the EPG3 annotation programme. All 7 annotation points were made for tokens where two places of articulation were achieved. That is, 7 annotations were made for /n#k/ realisations in careful or fast speech where the alveolar target was unassimilated. Figure 2.6 shows an example of a token thus annotated: *...ban cuts...*careful speech. Annotation points are made on the acoustic waveform as shown; only the relevant portion of the waveform is shown. It must be noted that the order of annotation points as they appear in Figure 2.6 varies from token to token. For example, voicing can end before alveolar release.

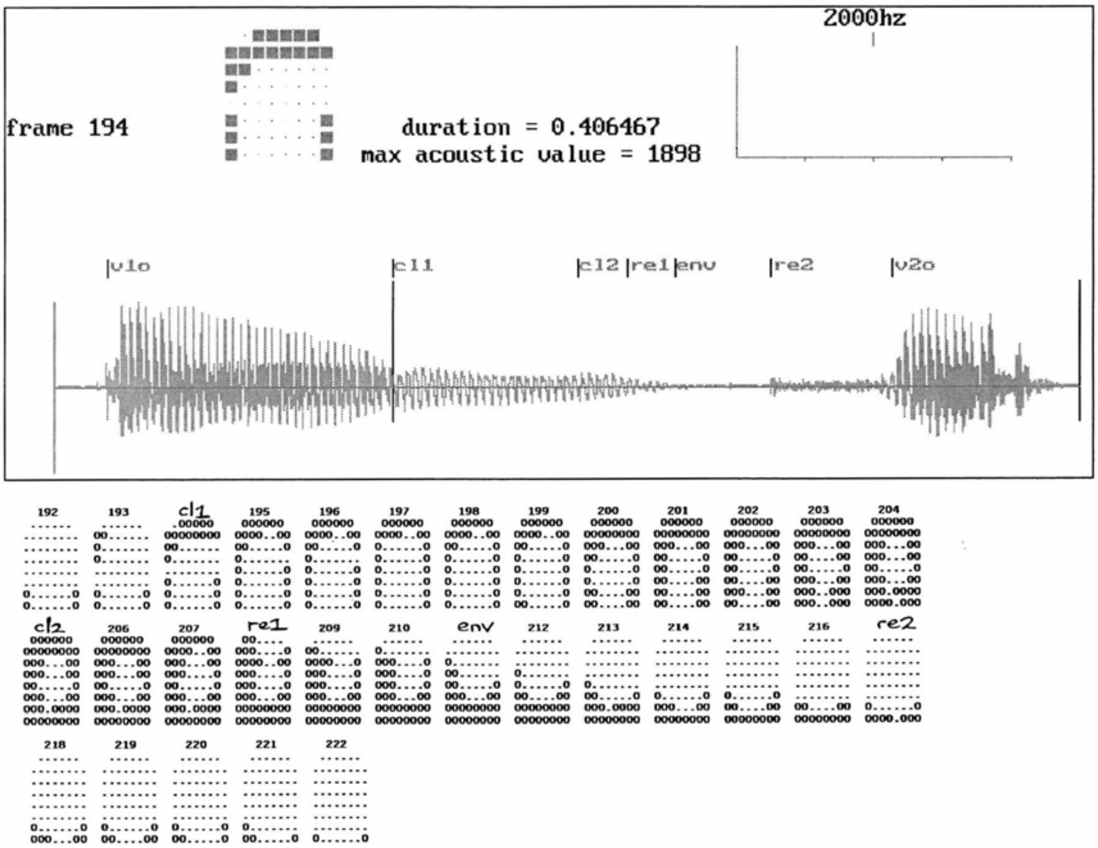


*Figure 2.6 EPG3 screen display of ...*ban cuts...* in careful speech showing annotated waveform (upper panel). The single EPG frame at the top shows the first frame of alveolar closure. The lower panel shows the EPG print-out for this token - annotation points corresponding to those on the waveform are marked – vowel onset annotation points are not shown*

Only 5 annotation points were made for sequences where a single place of articulation was achieved i.e. /ŋ#k/ *...bang comes...* in fast and careful conditions and /n#k/ *...ban cuts...* in the fast rate primarily, where an assimilation had taken place. Figure 2.7 shows an annotated /ŋ#k/ token in fast speech.

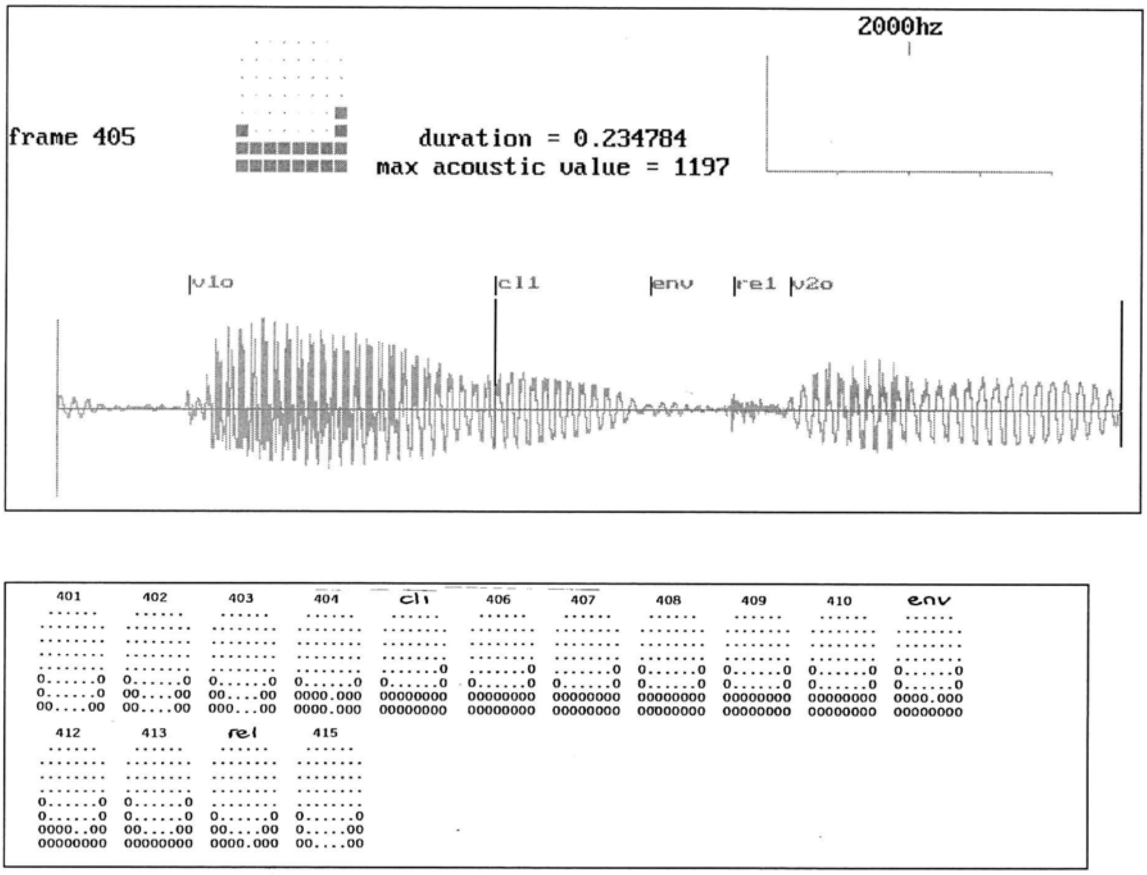


*Figure 2.7 EPG3 screen display of ...*bang comes... *in fast speech showing annotated waveform (upper panel). The single EPG frame at the top shows the first frame of velar closure. The lower panel shows the EPG print-out for this token - annotation points corresponding to those on the waveform are marked - vowel onset annotation points are not shown*

For target alveolar realizations, the annotation procedure allows for either the presence or absence of an alveolar only. That is, *partial* assimilations are not formally annotated. The criteria for annotation of a contact pattern as an alveolar stop closure is that there must be some mid sagittal contact in the alveolar region (criteria taken from Hardcastle, 1994). It is important to note that mid sagittal contact and not complete closure across the alveolar ridge is the classificatory criterion used here. This is because incomplete closure has in fact shown to be a common in non-assimilatory contexts in speech. Farnetani (1990) has amongst others shown that VCV consonantal weakening is a function of speech rate. She has shown using EPG that voiced stops in particular are reduced spatially by increase in speech rate. Figure 2.3 delimits the mid sagittal area. Care was taken in the case of 'borderline' tokens. Figure 2.8 illustrates a contact pattern for target /n/ which *was* annotated as an alveolar closure, even though not all electrodes in row 1 or 2 are contacted. Stop closure onset is at frame 253 and release is at frame 257. It is possible that absence of stop closure contact on the left hand side of the alveolar ridge is due to contact made on that side with the teeth. By contrast, Figure 2.9 is an illustration of a contact pattern for target /n/ which was *not* annotated as alveolar closure due to lack of mid sagittal contact. Both illustrations are from fast speech. The extensive lateral contact into the alveolar region (up to and including row 1) in frames 108-112, however, indicates that this is not a complete assimilation. Instead this token is considered a partial assimilation/residual alveolar because the tongue body has still made the supporting gesture in the absence of tongue tip/blade contact. A tongue body gesture is evident in Figures 2.8 and 2.9 but the crucial factor of mid sagittal contact separates them as different type of alveolar realisation. Following Nolan (1992), Barry (1985) and Kerswill (1985), there is a further factor which determines whether a token can be identified as a residual alveolar. If extensive side contact is characteristic of the same speaker's lexical /ŋ/ productions, then there is no justification for labelling it as a residual alveolar. This simply means that the particular speaker has fronted velar articulations.

```
  252       253       254       255       256       257       258       259       260       261       262       263
 ......    .....0    .....0    ......    ......    ......    ......    ......    ......    ......    ......    ......
......00  ....0000  ....0000  ....0000  ....0000  .....000  ......00  ......0.  ........  ........  ........  ........
.......0  0......0  0.....00  0.....00  0....000  0....000  0....000  0.....00  ......00  .......0  .......0  .......0
.......0  0......0  0......0  0......0  0.....00  0.....00  0.....00  0.....00  0.....00  0.....00  0......0  0......0
0......0  0......0  0......0  0.....00  0.....00  0.....00  0.....00  0.....00  0.....00  0.....00  0.....00  0.....00
0......0  0......0  00.....0  00....00  00....00  00....00  00....00  00....00  000...00  000...00  00....00  00....00
0......0  00....00  00....00  00....00  00....00  000...00  000...00  000...00  000...00  000...00  000...00  000...00
0......0  00....00  00....00  00...000  000..000  00000000  00000000  00000000  00000000  00000000  00000000  00000000
```

Figure 2.8 contact patterns for target /n#k/ fast speech – alveolar realisation shows mid sagittal contact and so is annotated as alveolar stop closure beginning at frame 253

```
  106       107       108       109       110       111       112       113       114       115       116       117       118
 ......    ......    0.....    0.....    0.....    0.....    0.....    ......    ......    ......    ......    ......    ......
........  0.......  0.......  0......0  00.....0  00.....0  00.....0  00.....0  0......0  0.......  ........  ........  ........
0......0  0......0  0......0  0......0  0......0  00.....0  00.....0  0......0  0......0  0......0  0......0  0......0  0.......
0......0  0......0  0......0  0......0  0......0  0......0  0......0  0......0  0......0  0......0  0......0  0......0  .......0
0......0  0......0  0......0  0......0  0.....00  0.....00  0.....00  0.....00  0.....00  0......0  0......0  0......0  0......0
0......0  0......0  00.....0  00....00  00....00  00....00  00....00  00....00  00....00  00....00  00....00  00.....0  0......0
0......0  0......0  00....00  00....00  00....00  00...000  00...000  00...000  00...000  00...000  00....00  00....00  00....00
00......  00....0.  00....00  000..000  000..000  0000.000  00000000  00000000  00000000  00000000  00000000  00000000  0000.000
```

Figure 2.9 contact patterns for target /n#k/ fast speech – alveolar realisation shows an absence of mid sagittal contact and so no annotation of alveolar stop closure is made.

A database was designed in Microsoft Access (version 2.0) which stores the imported annotation information (the tongue-palate contact for each annotated EPG frame and its time value). From this database, spatial and timing indices could be calculated.

2.1.5.2 Data analysis

This section describes the various data reduction procedures performed on the EPG and acoustic data and the timing measures that were of interest. Of concern was both inter and intra-subject variability. Several qualitative graphical displays were generated from the data. The spatial analysis will be described first, followed by the timing analysis.

2.1.5.2.1 Spatial analysis

'Contact totals' displays were generated by the EPG software. The number of activated electrodes in three areas of the palate as a function of time is calculated and displayed. Their purpose is to show spatial trends and to a lesser extent timing trends in alveolar to velar sequences produced by individual subjects. Because of the interest in observing inter-speaker variability, individual subjects' 10 repetitions of a single experimental sequence were shown on a single contact totals display. Figure 2.10 shows a sample display for one subject's 10 repetitions of /n#k/. The display shows the amount of contact in three articulatory regions (refer to Figure 2.3) from the onset of the vowel /a/ up to beginning of /s/ in target *...ban cuts...* Time is indicated in EPG frames on the x-axis and the y-axis shows number of electrodes contacted. An impression can be gained of the timing of maximum constriction in the alveolar relative to maximum constriction in the velar region, as can variability in the amount of alveolar contact across repetitions. Location of contacted electrodes within the three regions is not indicated.

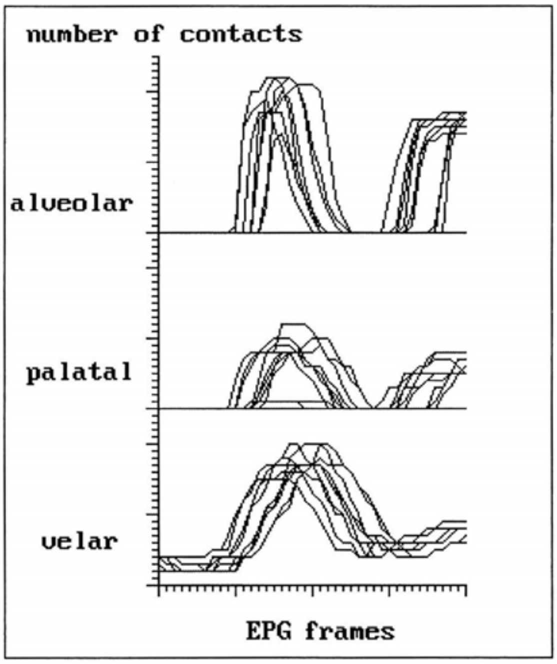


Figure 2.10 sample contact 'totals' display showing amount of contact in three articulatory regions as a function of time from the onset of the vowel /a/ up to beginning of /s/ in ...ban cuts... *(fast speech). Time is indicated in EPG frames on the x-axis, y-axis shows number of electrodes contacted in each region*

One other graphical display of spatial variability over 10 repetitions for individual speakers was generated, this time for particular annotation points. The access database was programmed to generate a single EPG frame (a 'prototypical' frame) for any annotation point. An example is given in Figure 2.11 below. A prototypical EPG frame for the annotation point *cl1*, which in this case is alveolar closure, is shown for one speaker (careful speech). Each individual square on the palate represents an electrode. The shading indicates the percentage frequency of contact for the electrode over the 10 repetitions (the scale is shown to the right in Figure 2.11), whereas the number shown within each square indicates the actual number of times the electrode was contacted. Thus, the black or white electrodes (the two extremes) indicate that there was no variability in contact over 10 repetitions.

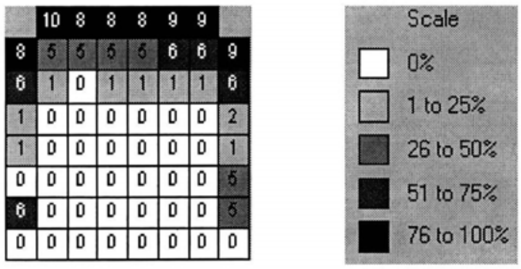


Figure 2.11 'prototypical' EPG display showing number of times electrodes were contacted over 10 successive repetitions of /n#k/, careful speech, for one speaker. Annotation point shown is onset of alveolar closure. Shading indicates percentage frequency of contact (scale to the right)

2.1.5.2.2 Timing analysis

The other aspect of the data that was assessed was timing of the phonetic events for /n#k/. 'Timing bars' using the software Origin 5.0 were constructed to linearly represent the coordination and duration in milliseconds of phonetic events for all tokens of /n#k/ in careful speech and fast speech. The phonetic events are EPG-defined articulatory events (corresponding to the data annotation points described in section 2.1.5.1) and the acoustically defined onset and offset of voicing. The representations of EPG events are thus of actual articulatory traces and not underlying gestures. Figure 2.12 shows a single labelled timing bar display for a non-assimilated token produced in careful speech. Time is shown in milliseconds on the x-axis. The grey timing bar at the bottom represents the period of voicing (the start point is 0ms) up to the onset of the voiceless phase of oral closure. The grey bar thus represents the duration of the vowel /a/ and the nasal consonant combined. The blue bar the top of the display represents the alveolar closure from its onset up to its release. The red bar represents the velar stop from onset up to release. The blue and red bars represent the lingual events involved in the production of /n#k/ and are generated from the tongue-palate contact data, whereas the grey bar represents the laryngeal events measured from the waveform. Where an assimilation has taken place, only two bars will appear on the display, one being the voicing bar and the other the red velar stop closure bar. Residual alveolars will not be represented with a blue bar, since their onset and offsets are difficult to identify, but the display of the token they appear in will be marked "residual alveolar". The occurrence of assimilations across speakers and their durational properties can be clearly seen in these displays. Figure 2.13 shows a set of 10 timing bars for one speaker's /n#k/ repetitions in careful speech as they appear in Chapter Three 'Results'.

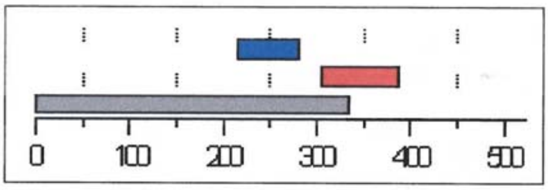


Figure 2.12 sample timing bar display showing coordination of laryngeal and supralaryngeal targets for one careful speech repetition of /n#k/ blue bar=alveolar stop, red bar=velar stop, grey bar=voicing

A number of measurements were made from the annotated data, such as those relating to durations of oral stops and indices of articulatory overlap in the adjacent stop closures. Graphs were created in the spreadsheet software Unistat. In order to investigate the temporal coordination of two adjacent stop closures in unassimilated /n#k/ sequences, two measures were made. Firstly, temporal latency (the time taken for velar closure to begin following alveolar closure) was considered an important indicator of intergestural control following Hardcastle and Roach (1979) and Byrd (1996). This was calculated as the interval in time between annotation points *cl1* and *cl2*. It may be hypothesised that the interval between these events becomes shorter as speech rate increases until the interval is so short that an assimilation takes place. This would mean that assimilation has a mechanical basis. Another important index of intergestural control for adjacent stops is duration of articulatory overlap, calculated as the interval between velar closure (*cl2*) and alveolar release (*re1*). If no simultaneous alveolar and velar closure is present, then this duration will be a minus value. Statistical analyses of this data will show whether these indices are significantly different in fast speech compared to careful speech and thus whether at faster rates of speech stop sequence production is reorganised.

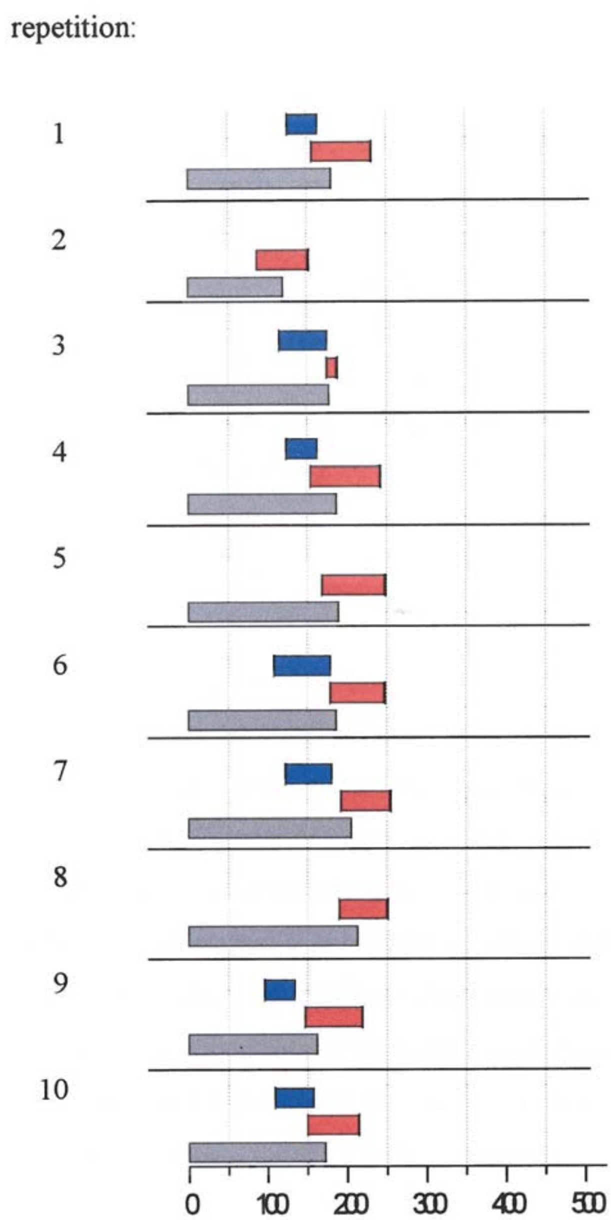


Figure 2.13 sample set of timing bars showing coordination and duration in ms of phonetic events for 10 repetitions of /n#k/ fast speech for one speaker. Blue bars represent the alveolar stop, red bars represent the velar stop and the grey bar at the bottom represents the period of voicing (starting at 0ms). The consonant bars are produced from EPG patterns and the voicing bar is produced from the waveform.

The other important calculation involves the duration of velar stop /ŋ/. This is measured as the duration between articulatory onset of the velar nasal up to the end of voicing which signals the onset of the voiceless phase of oral closure for /k/ (the end of voicing is here taken to signal the nasal-oral transition although it is acknowledged that there are no instrumental means of confirming this using velum movement tracking). This is an important measurement in that a phonetic comparison can be made between /ŋ/ derived from assimilated /n#k/ (where assimilation has been judged to be complete) and lexical /ŋ/, as in the control sequence /ŋ#k/. If it turns out that the duration of the derived and the lexical /ŋ/ are not statistically significantly different then it can be surmised that the contrast between target /n/ and target /ŋ/ has been collapsed (notwithstanding some other more elusive and subtle phonetic difference).

In order to control for the variation in speech rate between tokens being compared (variation either between the two speaking conditions or between individual tokens within either of the conditions), measurements were expressed as percentages of a longer stretch of speech as well as absolute values. This longer stretch of speech was taken from the onset of the vowel /a/ up to the end of frication for /s/ in *...ban cuts...*or *...bang comes...*This stretch of speech was also used as a measure of speech rate in order to establish that all subjects did in fact speed up their rate when producing test items in the fast/casual part of the experiment. Furthermore, if assimilations occurred in either of the experimental speaking conditions then a method of ascertaining if this was caused by an increase in rate relative to non-assimilated tokens would be needed.

2.2 COMBINED EPG/EMA EXPERIMENT

This section describes the methodology for the follow-up combined EPG/EMA study in the following order:

1. EMA instrumentation and set up
2. Preparations for recording
3. Post session data management
4. Data analysis

Two subjects were recorded, for reasons which will become clear following Chapter Three: 'Results'. The stimuli and elicitation procedure were identical to that of the EPG-only experiment described in section 2.1.

2.2.1 EMA instrumentation

Kinematic data was recorded by the simultaneous use of Electromagnetic Articulography (EMA), the Carstens *Articulograph Medizinelektronik AG100* (www.articulograph.de) and the Reading EPG2 system. This technique tracks articulatory movements by means of small electromagnetic transducers attached to the articulators in the mid-sagittal plane (see Hoole, 1993, Perkell et al, 1992 for more details). During data acquisition, three transmitter coils fixed on a helmet worn by subjects produce an alternating magnetic field at different frequencies. An equilateral triangle is formed by this field whose sides limit the measurement plane of the Articulograph AG100. Figure 2.14 shows a subject prior to an EPG/EMA recording session with the helmet in position which houses the three transmitter coils. The alternating magnetic field induces an alternating current in the transducers attached to the articulators (sensor coils) and from this the distances of each sensor coil from the three transmitters can be obtained. The XY coordinates are calculated and thus the positions of the sensors can be measured, stored and displayed for data analysis.

The EMA signal was synchronised with the EPG signal by the use of serial port communication, see Wrench (in press) for more details.

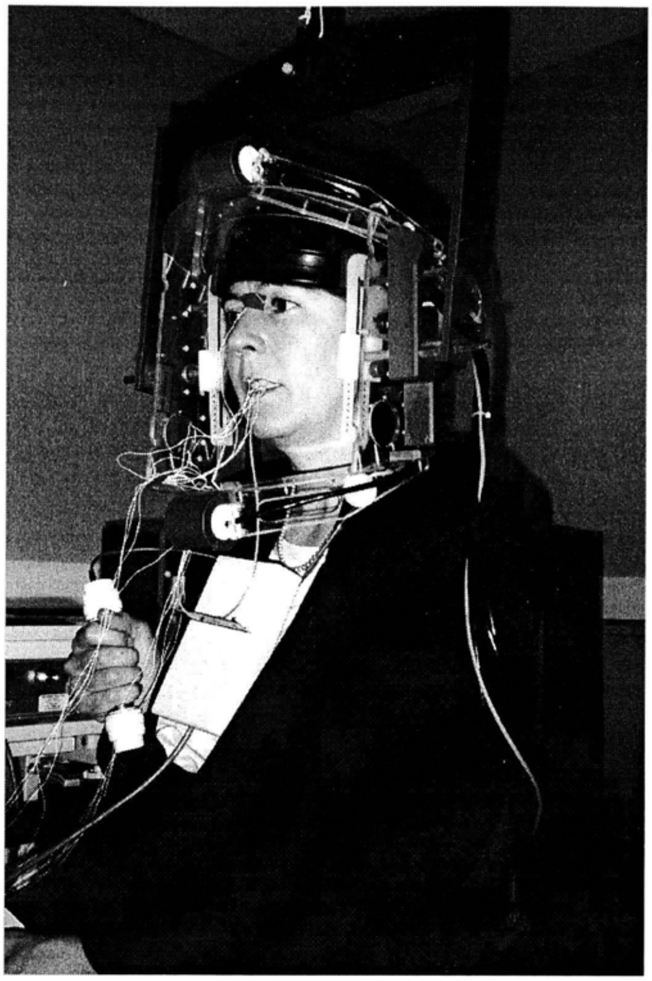


Figure 2.14 subject ready to begin an EPG/EMA recording session

2.2.2 Preparations for recording

Prior to each EPG/EMA recording the instrumentation was calibrated using the Carstens auto-calibration system. Also before each recording sensor coils were cleaned, sterilised and coated in latex which dries to form a strong rubber coating over the sensor coil and neck leading to the wires. Two hours before a recording, the Articulograph is switched on in order for it to warm up to a stable operating temperature.

2.2.2.1 Sensor coil positioning

Subjects were seated in a purpose-built sound-damped studio. A total number of 8 sensors were attached to the articulators *on the midline*. Table 2.3 shows the order of placement of coils and the position of the tongue coils and Figure 2.15 shows the placement of these sensors pictorially. Some sensor coils were used for the purposes of post-processing such as head movement correction (coils 1 and 8) and rotation (coil 9 'occlusal plane') to allow data normalisation across subjects and between tokens from the same recording session (these procedures will be described in more detail in section 2.2.3 below). Coil 9 was also used to obtain a 'palate trace'. Palate traces were obtained for each subject for the purpose of a visual reference during data analysis for a particular session (see Figure 2.17 for an example). It allows coil trajectory displays to be oriented within the vocal tract, so that tongue tip trajectories are not mistaken for tongue dorsum trajectories. This is an approximate reference only. Palate traces were obtained from each subject after the recording was finished. A sensor coil is attached to the fleshy underside of the subject's thumb and then placed in the back of the mouth where the soft palate meets the posterior edge of the EPG palate. Then a recording is made while the sensor attached to the thumb is moved from the back of the palate to the front, up to the edge of the teeth, following the shape of the palate as closely as possible. Although coils were attached to the upper and lower lip (coils 3 and 4), as part of standard recording procedure in the Department for the purposes of tracking lip movement, analysis of lip movement data was not performed for this study.

Table 2.3 order of coil placement in preparation for EMA recording and position of tongue coils

Coil	Location
1	Upper incisor
2	Jaw (lower incisor)
3	Upper lip (vermilion border)
4	Lower lip (vermilion border)
5	Tongue tip (10mm back from the extended tongue tip)
6	Tongue middle (20-30mm back from coil 5)
7	Tongue dorsum (20-30mm back from coil 6)
8	Bridge of nose
9	Used for palate trace and occlusal plane

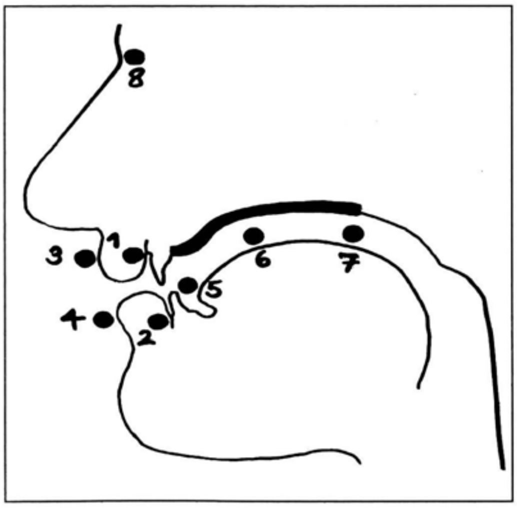


Figure 2.15 placement of sensor coils 1=upper incisor (reference), 2=lower incisor (jaw), 3=upper lip, 4=lower lip, 5=tongue tip, 6=tongue middle, 7=tongue dorsum, 8=bridge of nose (reference)

Coils were attached so that they lay within 3 mm either side of the mid-sagittal plane. Once coils have been attached, the helmet must be positioned. Care must be taken that once helmet has been positioned, the transmitter coils should also lie in the mid-sagittal plane.

2.2.3 Post session data management

Before data analysis could proceed the raw EMA data had to be post-processed. Firstly, the data was rotated. The method of rotation used here was to use a T-bar to record the occlusal plane of subjects (a technique suggested in Hoole, 1993). At the end of the recording session two sensor coils (coil 9 and one other which had already been used during the recording, typically a tongue coil, but not a reference coil) are attached to a flexible plastic T-bar at either end. The subject then places this in the mouth until it reaches the rear molars and bites down on it while a brief recording is made. All data for that particular session can then be rotated so that this occlusal plane is horizontal.

All other post-processing was carried out using the 'Tailor' software from Carstens. Data was filtered to smooth the reference coil values prior to head correction. Data was processed for head movement correction in the mid sagittal plane to ensure that the two reference coils register at the same geometrical co-ordinates. This compensates for any head movement, backwards, forwards, up and down relative to the helmet. See Hoole (1993) for an EMA data reliability study.

2.2.4 Data analysis

A set of programmes called *EMA tools* for the analysis of EPG/EMA data was written by Noel Nguyen (laboratory for Psycholinguistics, Geneva University) using the MATLAB programming language (Hoole and Nguyen, 1992; Nguyen, 1996). Some additional macros were written by Alan Wrench (Queen Margaret University College, Edinburgh). MATLAB integrates numerical analysis, matrix computation, signal processing and graphics in the same environment.

Figure 2.16 shows the 'Main Window' screen display in MATLAB which controls the entire interface. The window allows analysis, editing and annotation of acoustic and articulatory signals. Editing functions, such as zooming and annotating, are performed using two cursors which can be moved along the time axis with the mouse in the usual fashion. Individual articulatory and acoustic signals are plotted in sub-regions of the main window and are referred to as 'tracks'. Tracks selected for display in the Main Window for the purposes of this experiment, were of three different kinds:

1 the acoustic signal (waveform)
2 EMA parameters (the *x* or *y* coordinate of a sensor coil)
3 tangential velocity of a sensor coil (*x* and *y* velocity combined), derived by means of a mathematical transformation

A list of tracks selected is shown in Table 2.4:

Table 2.4 EMA tracks selected for data analysis in the order in which they appear on the Main Window and their corresponding labels - labels which include 'sm' are those signals which have been smoothed for easier identification of maxima and minima of peaks

trace type	**label on Main Window**
tongue tip *x* dimension	tt/x – sm
tongue tip *y* dimension	tt/y – sm
tongue tip tangential velocity	tt-tv
tongue dorsum *x* dimension	td/x – sm
tongue dorsum *y* dimension	td/y – sm
tongue dorsum tangential velocity	td-tv

From the Main Window additional displays of the data could be selected (see Nguyen, 1996, for complete list). The displays selected here were:

1 'EPG Pat'

a display of the EPG pattern that coincides with the left-hand cursor and the right-hand cursor respectively

2 'EPG Printout'

a display of *all* EPG patterns that fall within the position of the left and right cursor

3 'EMA Traj'

a display of the trajectory of each sensor coil in the mid-sagittal plane over the time interval delimited by the two cursors

Firstly, using the 'EPG Pat' and the 'EPG Printout' displays, the EPG data were checked to ascertain the number and type of alveolar assimilations that had occurred in each careful and fast speech for each subject.

Two types of measurement were made using the tracks display on the Main Window. The first involved identifying *x* and *y* articulatory coordinates for the tongue tip coil and the tongue dorsum coil at the onset of the articulatory beginning for the sequence (also used by Kühnert, 1993). The onset corresponds to the moment of maximum *tongue tip* displacement for all target /n#k/ sequences and the moment of maximum *tongue dorsum* displacement for all /ŋ#k/ sequences. Minimum tangential velocity was used to identify maximum displacement of the tongue tip coil in both dimensions. This was because maxima and minima of velocity peaks are more readily identifiable than displacement maxima and minima.

If tongue tip position for any assimilated /n#k/ (EPG-defined) token at this moment is higher than it is for any /ŋ#k/ tokens, then this may be identified as a residual alveolar gesture and thus complete assimilation has not taken place. By way of illustration, Figure 2.16 shows how the *x* and *y* position of the tongue tip coil at maximum displacement was measured for an /n#k/ token (the tongue dorsum position is not marked on this Figure). On this display, the waveform appears at the top with all the EMA traces ordered below (see Table 2.4 above for tracks selected for this experiment). In the case of *x* axis coil position, movement upwards on the display corresponds to actual horizontal movement backwards (i.e. movement towards the back of the mouth) and in the case of *y* axis coil position, movement upwards on the display corresponds to actual vertical tongue movement upwards. In the case of the velocity traces, upward movement indicates an increase in velocity. The left-hand solid black line in Figure 2.16 is placed at minimum tangential velocity for the tongue tip (third trace below the waveform *tt-tv*). The position of the right–hand solid line will be described below. Because all traces and the waveform are time aligned, position readings can be taken from the tt/x and tt/y position traces respectively. In Figure 2.16 a reading of 72.4mm is taken from tt/x and of 159.4mm from tt/y.

The second measurement taken from the Main Window was the time interval between maximum tongue tip displacement and maximum tongue dorsum displacement for the experimental sequences (max. displacement was, again, identified from min. tangential velocity). The purpose of this was to investigate the possibility that apparently complete /n#k/ sequences are subtly different from lexical /ŋ#k/ sequences in terms of timing of gestures. In Figure 2.16 the left-hand and right-hand cursors are placed at minimum tangential velocity for the tongue tip and the tongue dorsum respectively.

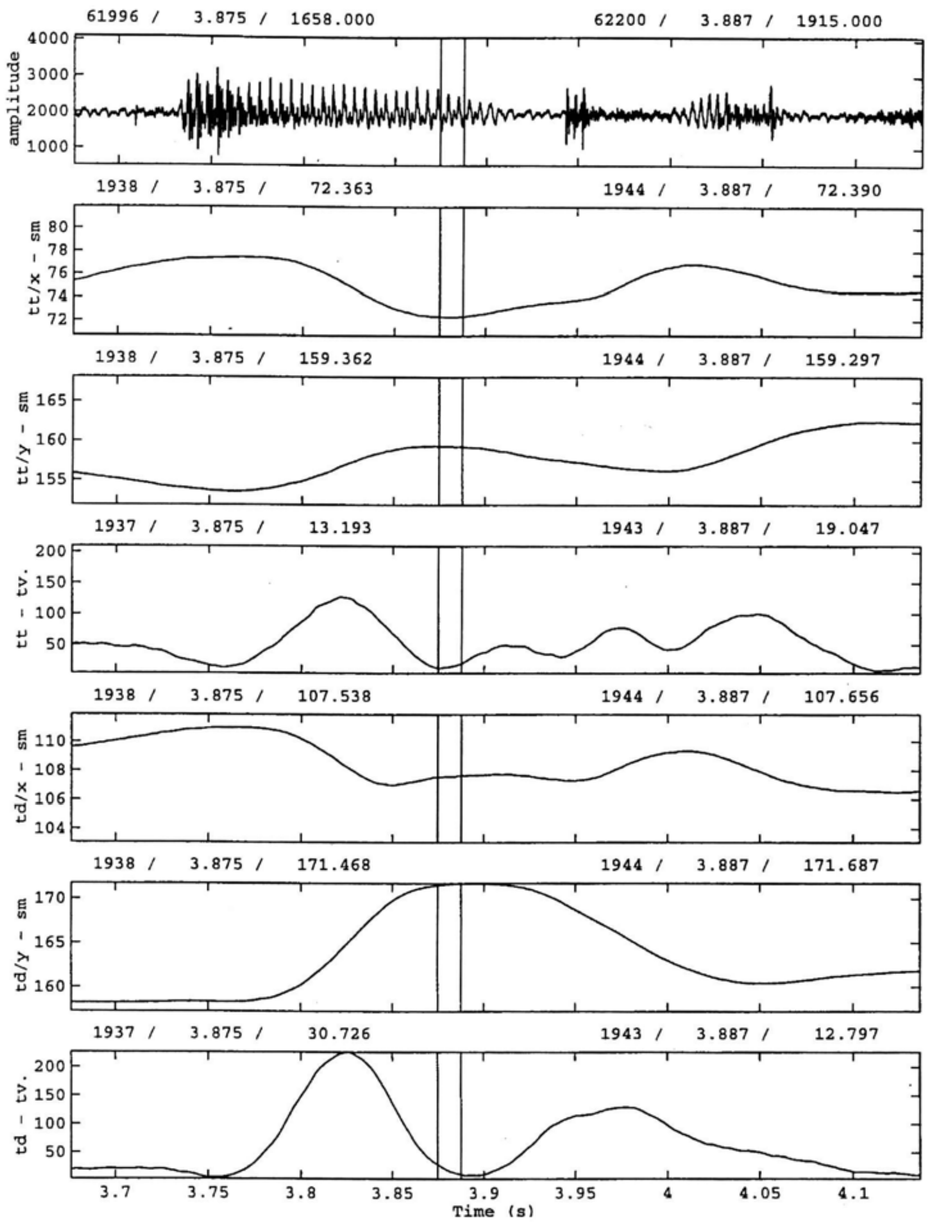


Figure 2.16 Main Window in MATLAB showing waveform and EMA traces for sequence /n#k/ fast speech left-hand cursor shows maximum displacement (minimum tangential velocity) for tongue tip coil, right-hand cursor shows the same point for tongue dorsum coil.

The only other display of the data used for analysis apart from EPG patterns and acoustic/EMA traces was the 'EMA traj' display. This was used as a visual reference indicating the spatial trajectory described by the three tongue coils over a particular time span. This time span was delimited by the left and right cursors on the Main Window. The left-hand cursor was placed at the middle of the pre-experimental sequence /a/ and the right-hand cursor was place at the middle of the post-experimental sequence /ʌ/, capturing the production of either target /n#k/ or /ŋ#k/. Figure 2.17 shows one such display for a single token.

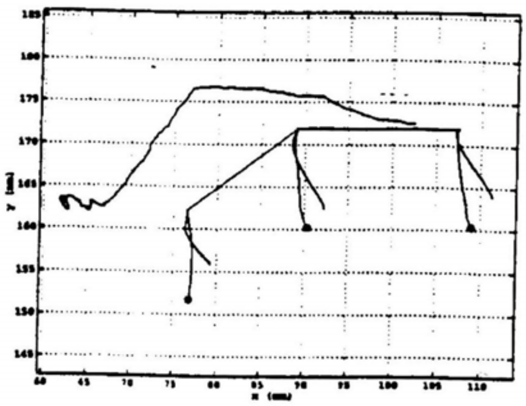


Figure 2.17 'EMA traj' display – showing spatial trajectory of the tongue tip (left-most coil), tongue middle and tongue dorsum coils for a production of /n#k/, the solid black line gives an impression of the overall tongue configuration at the articulatory beginning of a sequence.

CHAPTER THREE

EPG Results (Main study)

3.0 PRELIMINARY NOTES

All subjects produced the required number of repetitions of each experimental sentence in both speaking conditions. The only anomaly in articulatory ordering of the alveolar and velar gestures for /n#k/ was found in a single fast speech token produced by subject G. The EPG patterns for this (see Figure 3.4 (vii), repetition 9) indicate that the alveolar and velar stop closures are out of sequence. Unlike any other non-assimilation token in the entire database, alveolar contact builds up to maximum contact (frame 422) only *after* velar closure has been formed (frame 421). The normal sequence of events is for contact in the alveolar region to build up into a full stop closure before or at the same time as velar stop closure occurs. For the anomalous token, the alveolar closure and release phases are inaudible since they occur during the velar constriction. The resulting percept is of a complete alveolar assimilation and this token is annotated as a complete assimilation. There were no cases of *...ban cuts...*realised as [ban nʌts] although subject B produced three fast speech assimilated /n#k/ repetitions as [baŋ gʌts] , with no period of voicelessness for target /k/ (see Figure 3.16 (ii), fast speech repetitions 2, 4 and 5)

3.1 SPATIAL ASPECTS OF /n#k/

3.1.1. Overall occurrence of assimilation and measurement of speech rate

Total frequency of assimilations for /n#k/ fast and careful speech is shown in Table 3.1 and in Figure 3.1. For each speaking condition 100 tokens were yielded, 10 produced by each speaker.

Table 3.1 distribution of assimilations (all subjects combined) careful and fast speech

careful speech		fast speech	
non-assimilation	**assimilation**	**non-assimilation**	**assimilation**
96	**4**	**43**	**57**

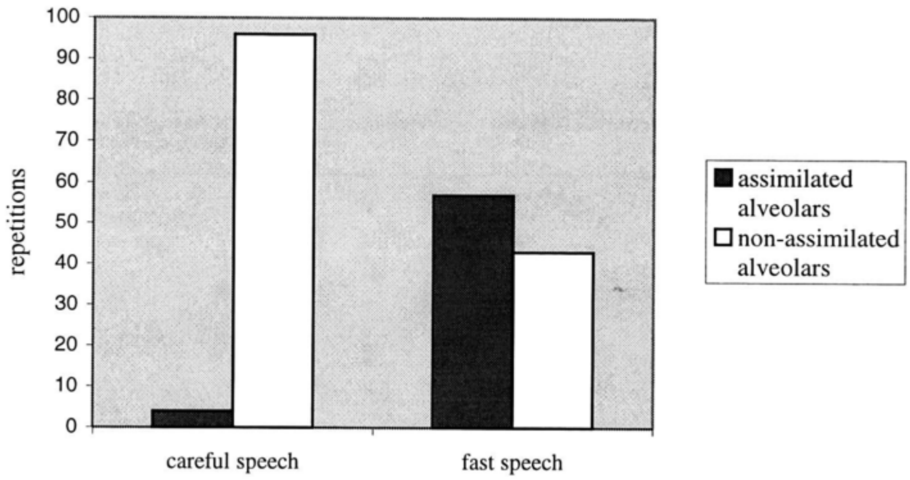


Figure 3.1 frequency of assimilation (all subjects) for /n#k/, careful and fast speech

As predicted, the change from careful to fast speech can be characterised by a considerable rise in the frequency of assimilations. Even though this is the case, the number of assimilations and non-assimilations for /n#k/ in fast speech is roughly the same, indicating that the correlation between assimilation and speech rate is not straightforward.

To establish that speakers did in fact speed up their speech on the instruction to produce the stimuli fast and casually, Figure 3.2 shows the speech rate of each production of the experimental sentence containing /n#k/ for all subjects in both the careful and the fast speech conditions. (It must be noted here that in this and subsequent scatterplots throughout Chapter Three, alphabetic subject labels on the x-axis will appear in lower case type.) Duration between the onset of /a/ in *ban* and the end of frication in /s/ in *cuts* was taken as an index of speech rate. This was chosen because it includes the consonantal sequence itself plus some contextualising speech. The line on the graph drawn at around 540ms separates the careful speech tokens from the fast speech tokens. Durational values for all fast speech tokens appear below this line. It is important to note here that there was no overlap between the duration for tokens from the different conditions. On no occasion was a fast speech token elicited from a subject measured to be slower in rate than a slow/careful token elicited from the same subject. Separate variance t-tests showed that all subjects' slow/careful tokens were significantly different in terms of speech rate than their fast tokens (a Shapiro-Wilk test confirmed that all subjects' samples had a normal distribution). This means that whenever the terms 'careful speech' and 'fast speech' are used in this dissertation it is done so on the understanding that fast speech refers to a set of tokens that were spoken at a measurably faster rate than the (slow and) careful speech tokens.

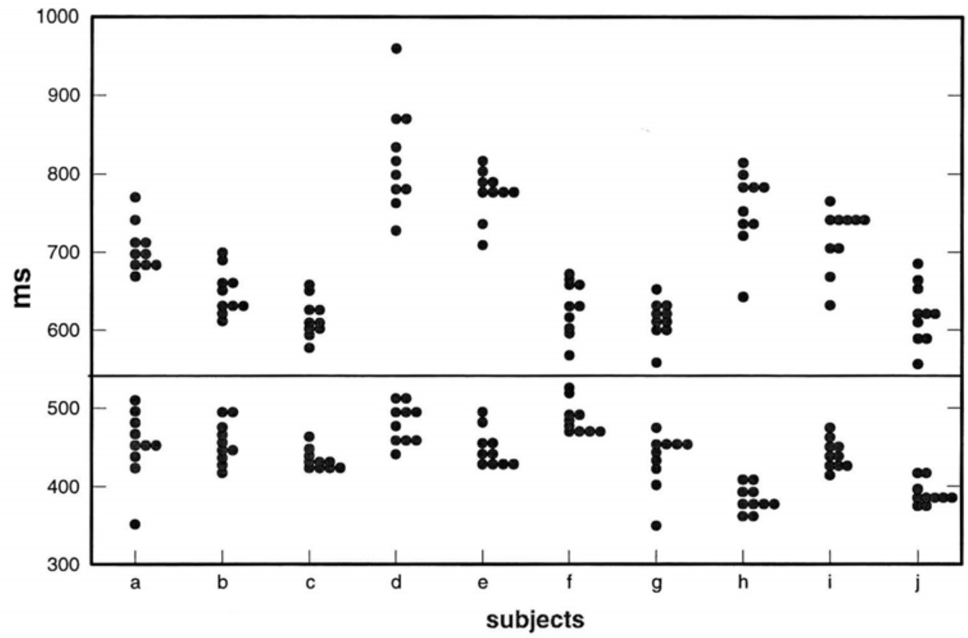


Figure 3.2: scatterplot showing rate of speech for /n#k/ fast speech tokens (below the line) and for careful speech tokens (above the line). Rate of speech was measured as time in ms between onset of /a/ in 'ban' *and the end of frication in /s/ of* 'cuts'. *Alphabetic labels for subjects appear in lowercase.*

Figure 3.2 shows that for some speakers, fast and careful speech are more distinct in terms of speech rate than for others. The two conditions are distinguished the most by subjects D, E and H, these speakers having some of the slowest careful speech tokens of all. Subjects F, however, has little difference in rate between the two conditions, although, as mentioned above, no overlap has taken place. The other point to make on the basis of Figure 3.2 is that there is more variation in rate between speakers in the careful condition than there is for the fast condition. This is possibly because there are physical limitations on producing very fast but intelligible speech beyond a certain rate whereas there are theoretically no such limitations on careful speech. There is no overlap at all between careful speech tokens from subject D and subject C or between subject E and subject G. Even though there are these quite different distributions for these subjects in careful speech, they both apply the same non-assimilation strategy when it comes to producing the experimental sequence (individual subject results will be presented in section 3.1.2 below). Table 3.2 shows the means and standard deviations of speech rate in each speaking condition across all subjects as a whole and below this, for individual subjects.

Table 3.2 means and standard deviations of speech rate in each speaking condition (100 repetitions ('cases') for each condition), subjects combined (top table) – bottom table shows means and standard deviations for each condition subject by subject

Rate:	Cases	Mean (ms)	Standard Deviation
careful	100	695.1	80.9
fast	100	445.7	40.8

subject × careful speech	Cases	Mean (ms)	Standard Deviation
A	10	712.9	27.3
B	10	653.5	26.9
C	10	619.6	23.9
D	10	826.3	63.9
E	10	779.5	30.2
F	10	632.9	34.5
G	10	617.6	25.8
H	10	759.3	49.4
I	10	725.0	40.5
J	10	624.4	36.3
subject × fast speech			
A	10	460.5	45.5
B	10	461.9	26.6
C	10	436.9	13.7
D	10	486.4	26.3
E	10	454.5	23.7
F	10	488.5	20.6
G	10	437.5	38.4
H	10	388.4	19.1
I	10	447.6	20.3
J	10	395.1	17.1

Number of assimilations for individual speakers is shown in Table 3.5 for both speech rate conditions. There were very few assimilations in careful speech. Subject G produced 1 and subject H produced 3 (these will be reported in more detail in section 3.1.3). These subjects always assimilated the alveolar stop in fast speech. Figure 3.3 (below) shows distribution of assimilations and non-assimilations for fast speech only. Subjects are arranged on the graph from left to right on the x-axis according to number of fast speech assimilations (assimilations are shown in black). Subjects fall into three basic groups in terms of frequency of assimilation in fast speech. Subjects E and F never assimilated; subjects A, B, C and D varied between assimilation and non-assimilation and subjects G, H, I and J always assimilated.

Table 3.5 frequency of assimilations for individual speakers, /n#k/ careful and fast speech

	careful		fast	
subject	**non-assimilation**	**assimilation**	**non-assimilation**	**assimilation**
A	**10**	**-**	**7**	**3**
B	**10**	**-**	**4**	**6**
C	**10**	**-**	**4**	**6**
D	**10**	**-**	**8**	**2**
E	**10**	**-**	**10**	**-**
F	**10**	**-**	**10**	**-**
G	**9**	**1**	**-**	**10**
H	**7**	**3**	**-**	**10**
I	**10**	**-**	**-**	**10**
J	**10**	**-**	**-**	**10**

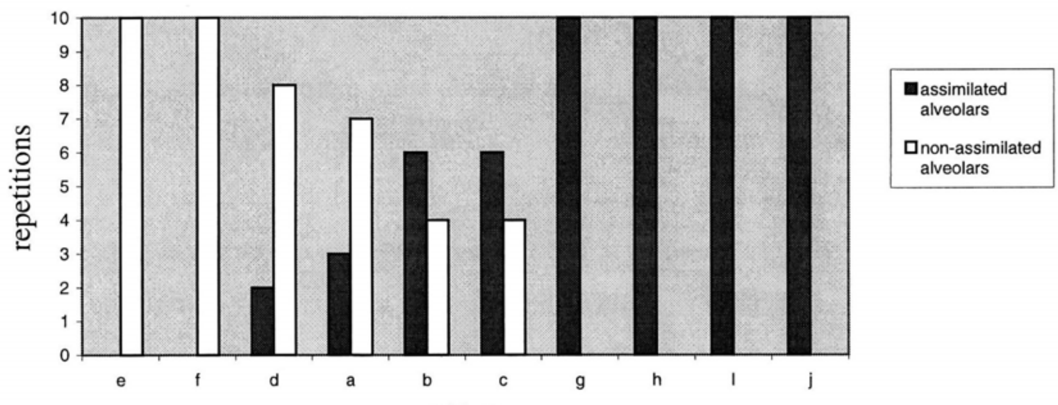


Figure 3.3 frequency of assimilations over 10 repetitions of /n#k/ fast speech for individual speakers

These results are based on the criteria for the data annotations points outlined in section 2.1.5.1 of Chapter Two. That is, tokens with no mid-sagittal contact on the alveolar ridge were classified as assimilations and no distinction is made, at this stage in reporting of the results, between degrees of assimilation such as ‘complete’ or ‘partial’. Occurrence of partial assimilations or ‘residual alveolar articulations’, are reported below.

3.1.2 Fast speech EPG data

Raw EPG patterns for fast speech /n#k/ tongue - palate contact are shown in Figure 3.4 (i)-(x) for all 10 subjects. Each single page shows 10 numbered repetitions produced by a particular subject. Each line shows a single repetition capturing the onset of alveolar closure, or velar closure if assimilation has taken place, and the release of velar closure for /k/. As described in Chapter Two section 2.1.5.2.1, the ‘contact totals’ displays for each subject on Figure 3.5 summarise, for all 10 repetitions, tongue - palate contact in the alveolar region from the start of the vowel in *...ban cuts...*approximately up to and including the /s/. That is, for each subject all ‘amount-of-contact’ curves (one curve = a single token) are superimposed onto one display in order to show general spatial trends and variability. Time in EPG frames is represented horizontally (frames sample every 10 ms). *Location* of contacted electrodes is not indicated. For the purposes of these displays, the palate is divided into three regions, the alveolar region (first three rows), the palatal region (fourth and fifth rows) and the velar region (rows six, seven and eight). The contact totals displays and particularly the raw EPG data will be referred to in detail for the report on subjects’ assimilation strategies which follows.

We are now in a position to look in more detail at assimilation strategies in fast speech. Results show that subjects fall into three broad categories: subjects who never assimilate (numbering 2), subjects who vary between non-assimilation and assimilation (4) and subjects who always assimilate (4). Within each section a descriptive label will be proposed for each group summarising the assimilation strategy used (this short-hand label will be used throughout this Chapter and also in Chapter Five: ‘Discussion’). Firstly, the data for the four subjects who always assimilate the alveolar will be considered; secondly the data for the four subjects who vary between non-assimilation and assimilation, and thirdly the remaining subjects who never assimilate in fast speech will be considered.

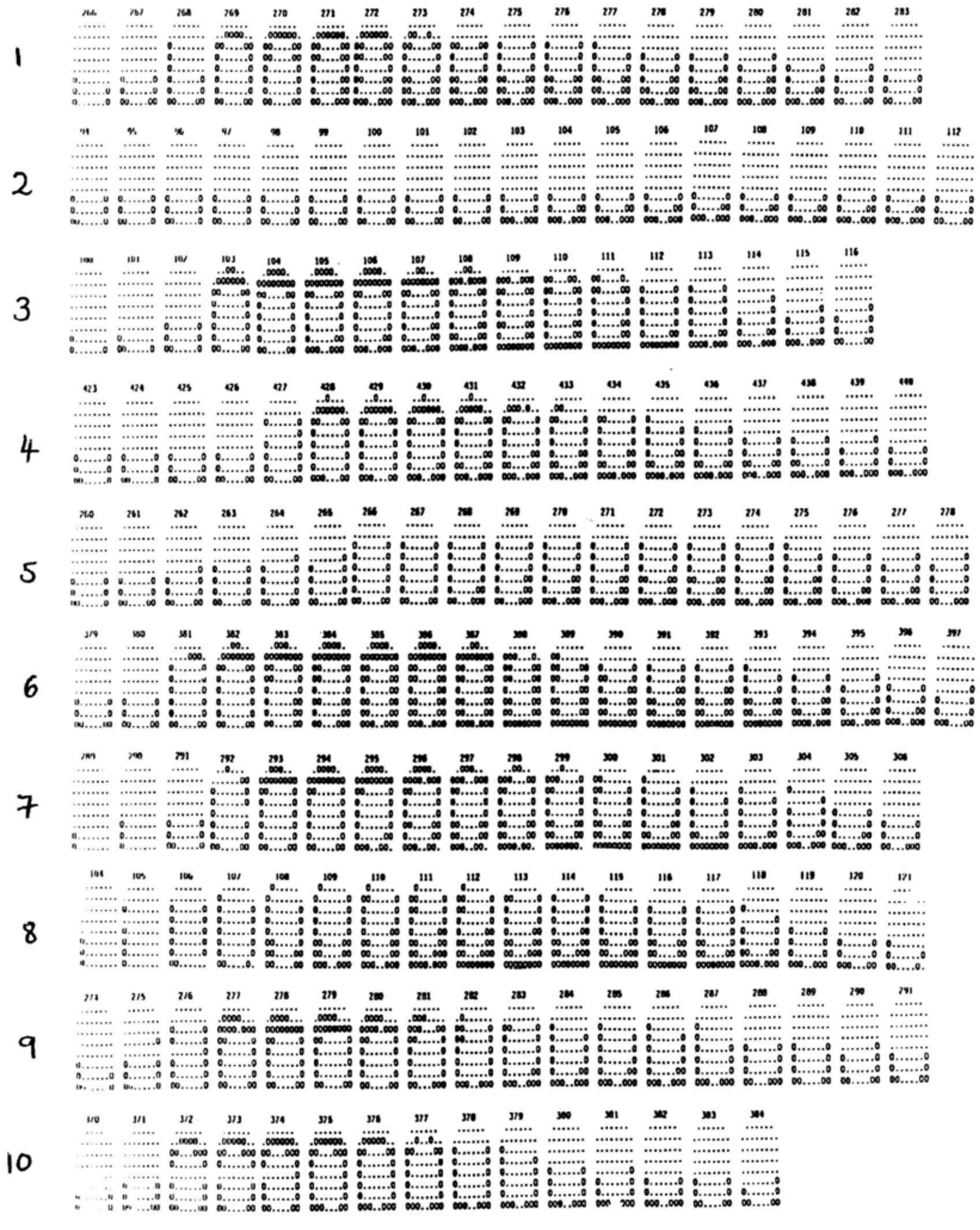


*Figure 3.4 (i): **subject A** EPG patterns for 10 fast speech /n#k/ repetitions - each numbered line shows a single repetition capturing the onset of alveolar closure, or velar closure if assimilation has taken place, up to and including the release of velar closure for /k/.*

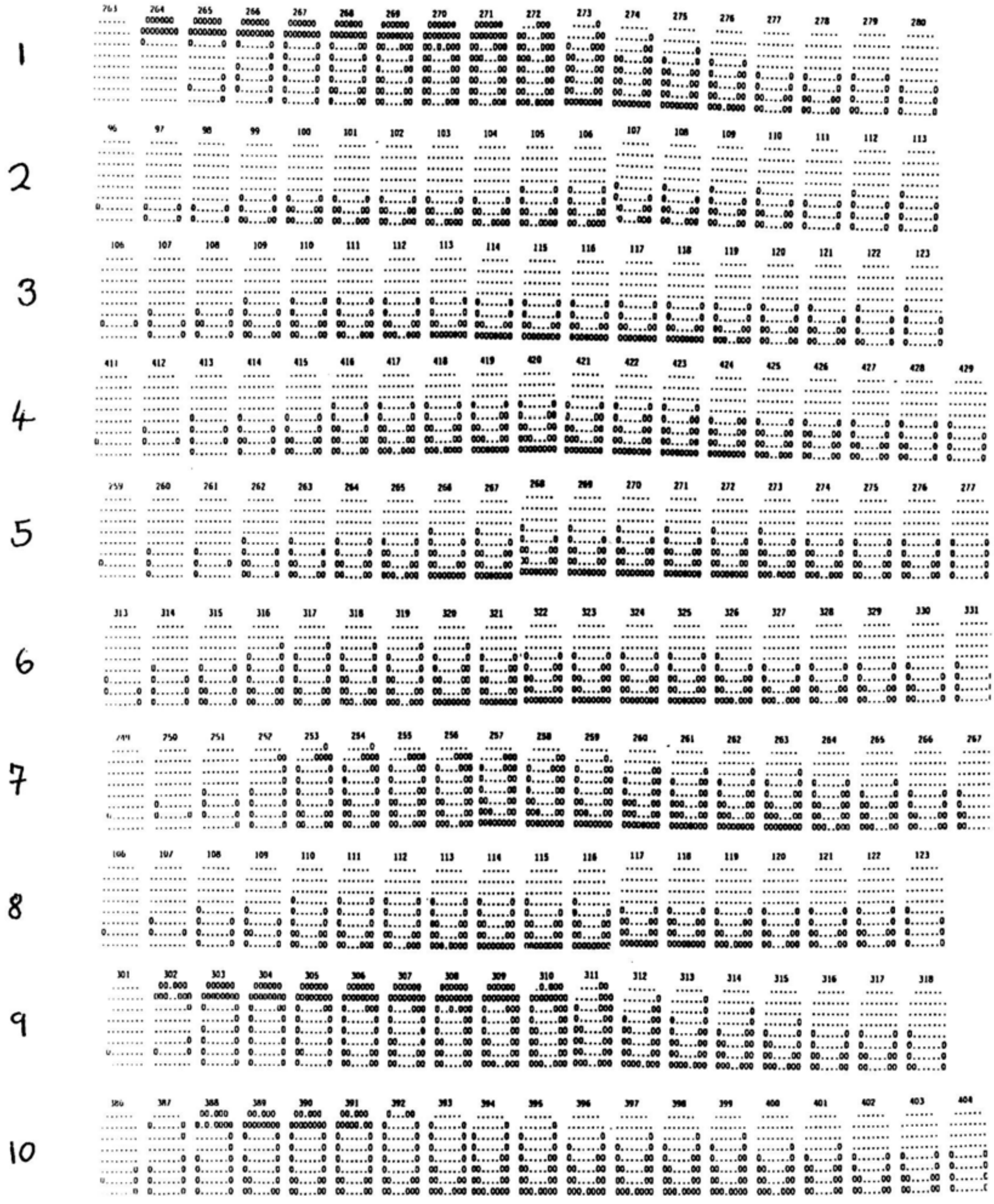


*Figure 3.4 (ii): **subject B** EPG patterns for 10 fast speech /n#k/ repetitions*

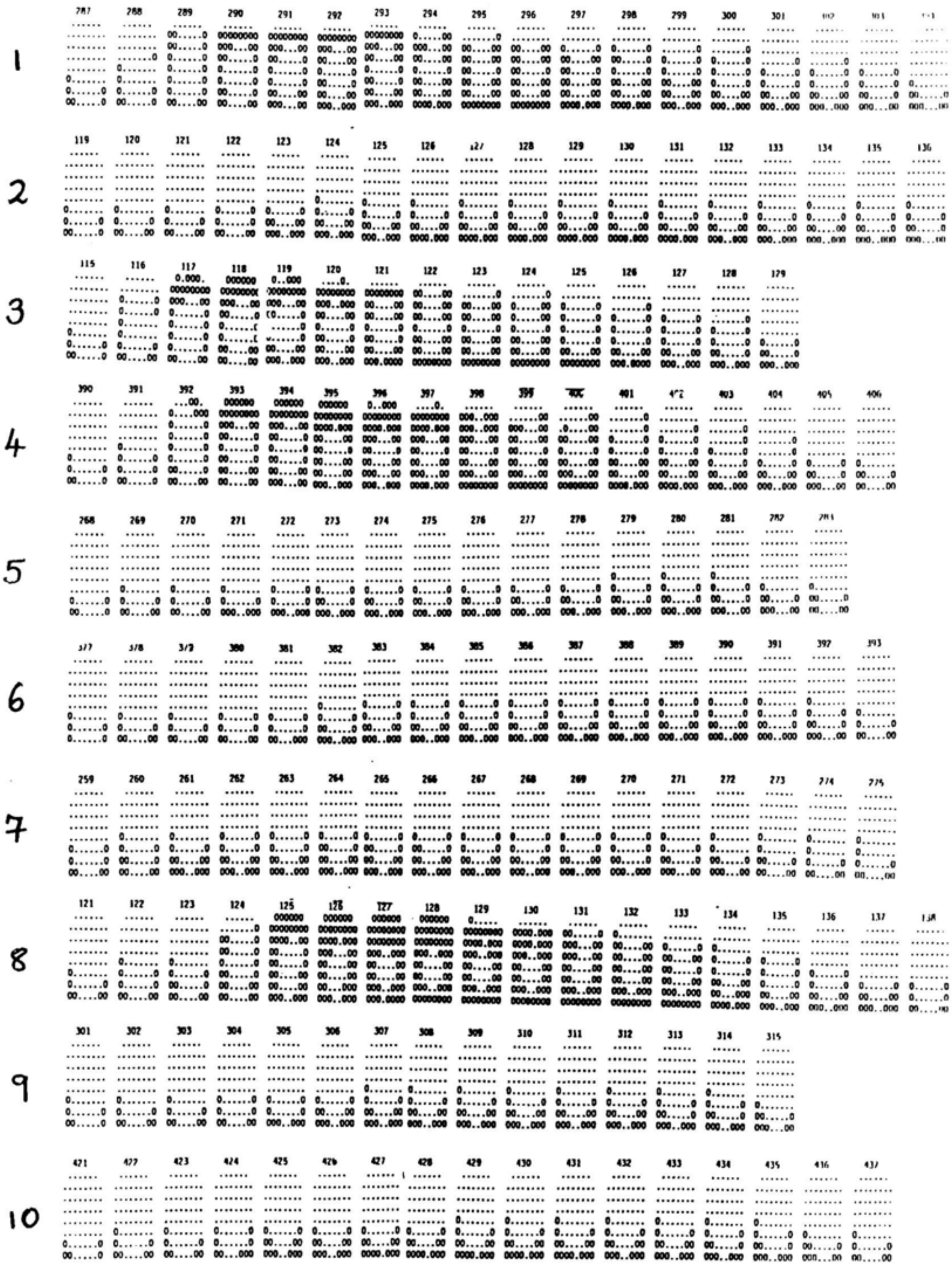


*Figure 3.4 (iii): **subject C** EPG patterns for 10 fast speech /n#k/ repetitions*

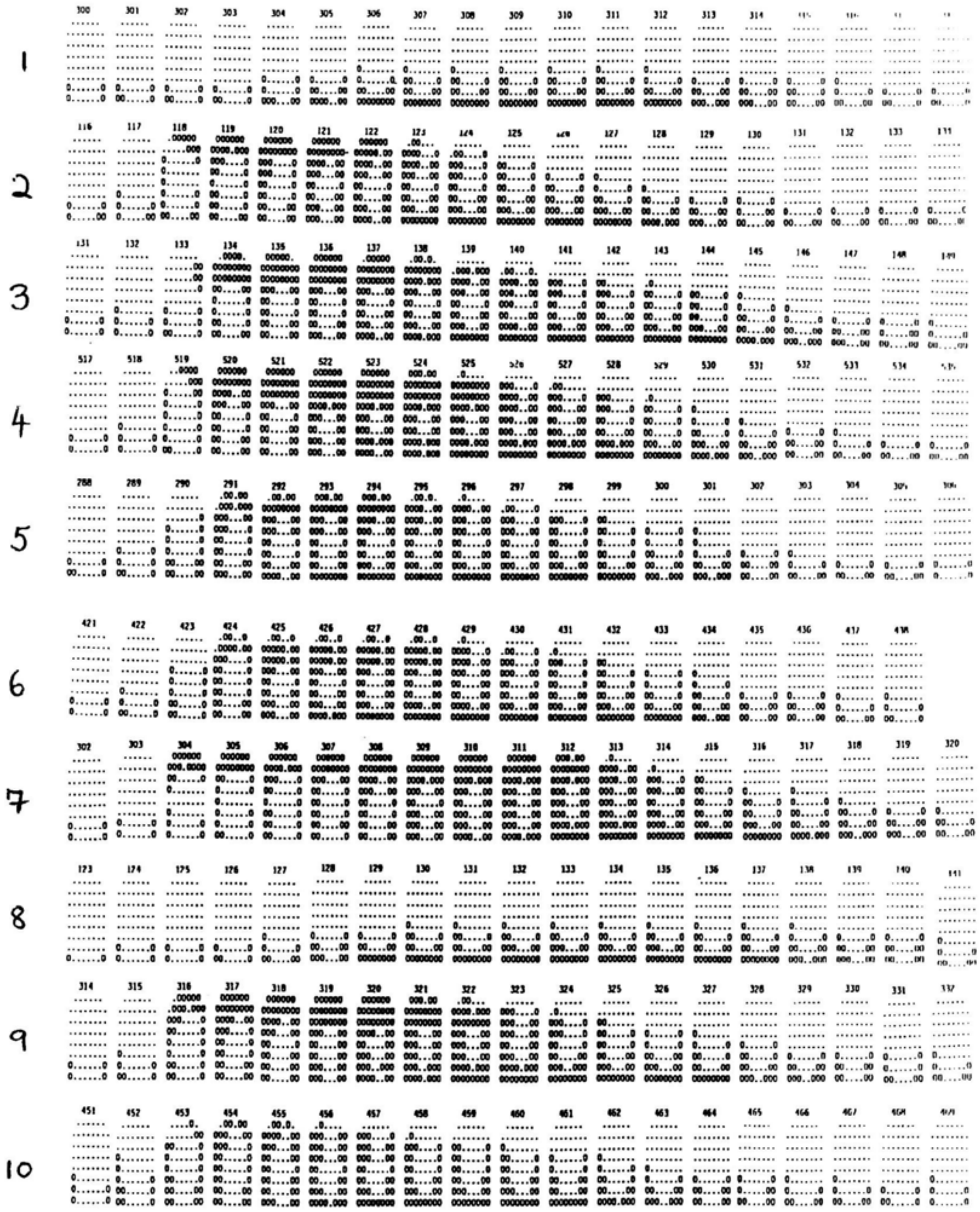


*Figure 3.4 (iv): **subject D** EPG patterns for 10 fast speech /n#k/ repetitions*

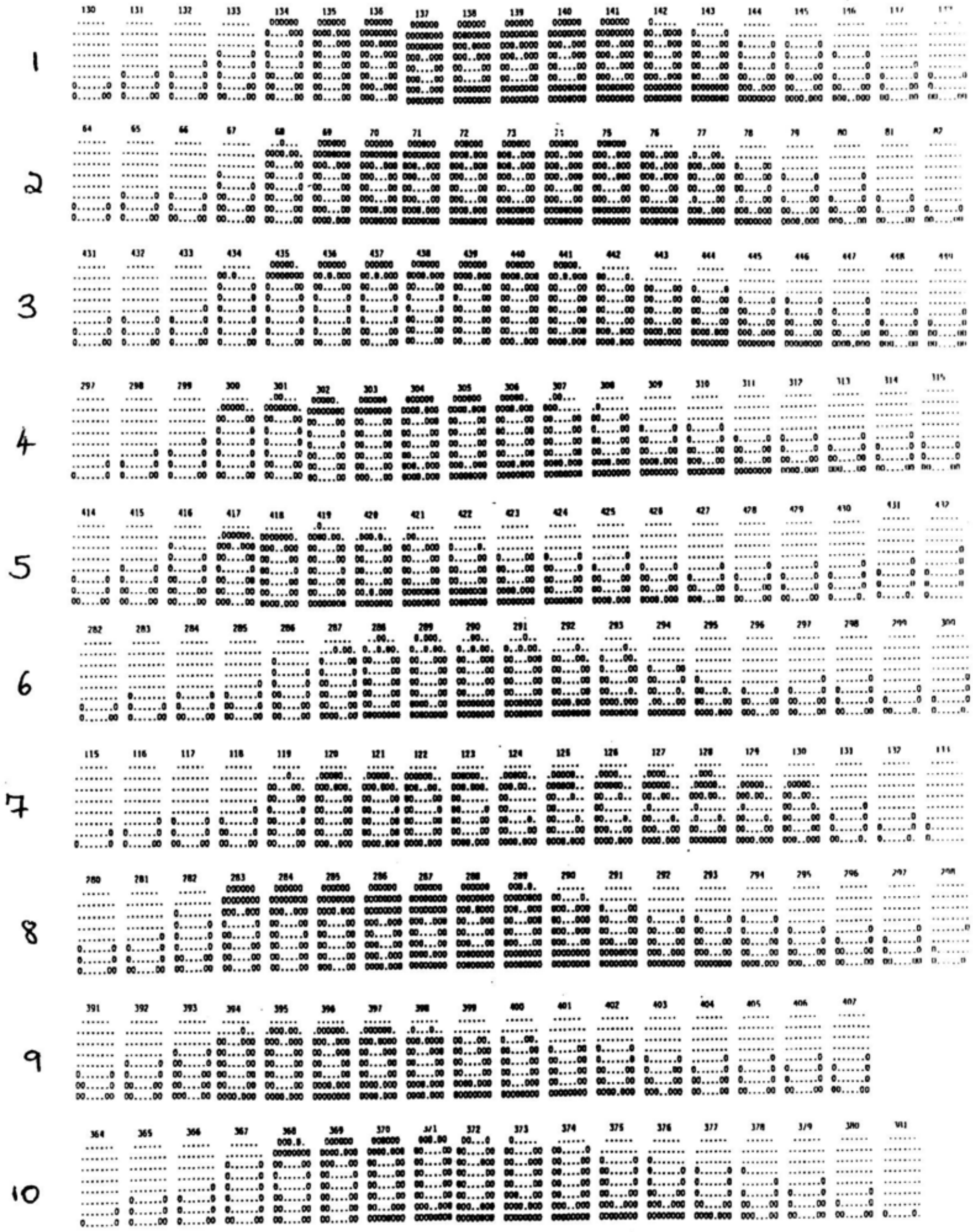


*Figure 3.4 (v): **subject E** EPG patterns for 10 fast speech /n#k/ repetitions*

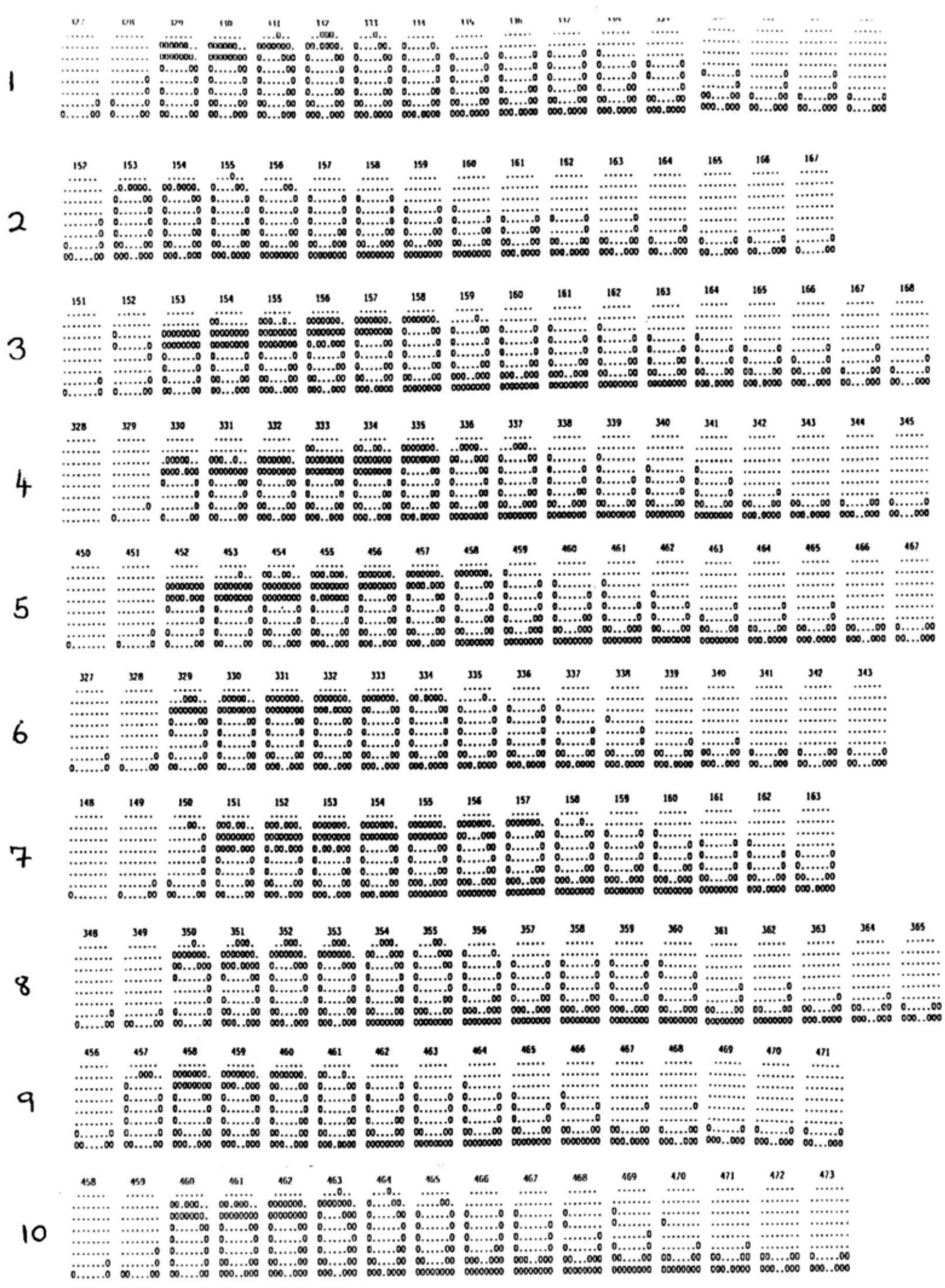


*Figure 3.4 (vi): **subject F** EPG patterns for 10 fast speech /n#k/ repetitions*

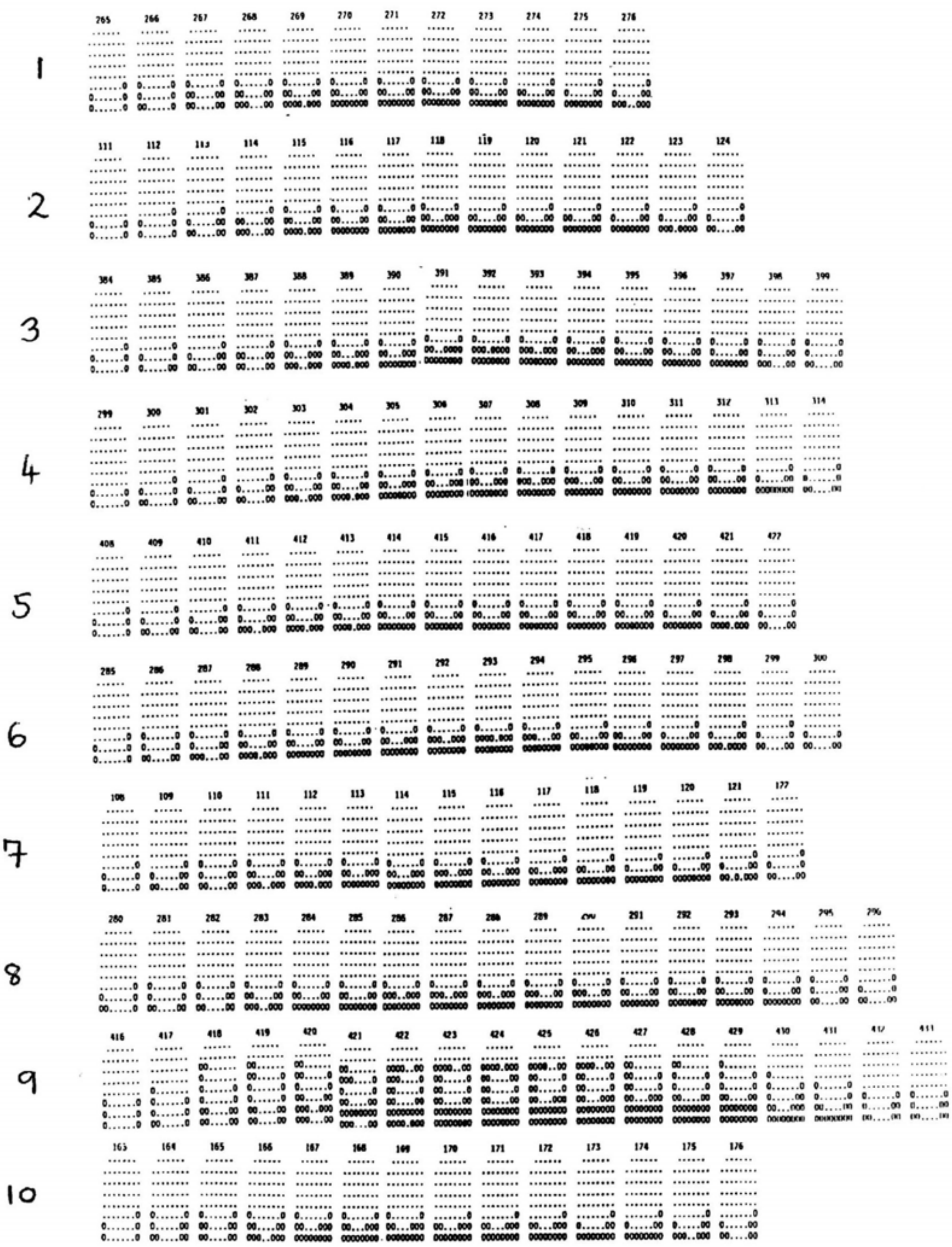


*Figure 3.4 (vii): **subject G** EPG patterns for 10 fast speech /n#k/ repetitions*

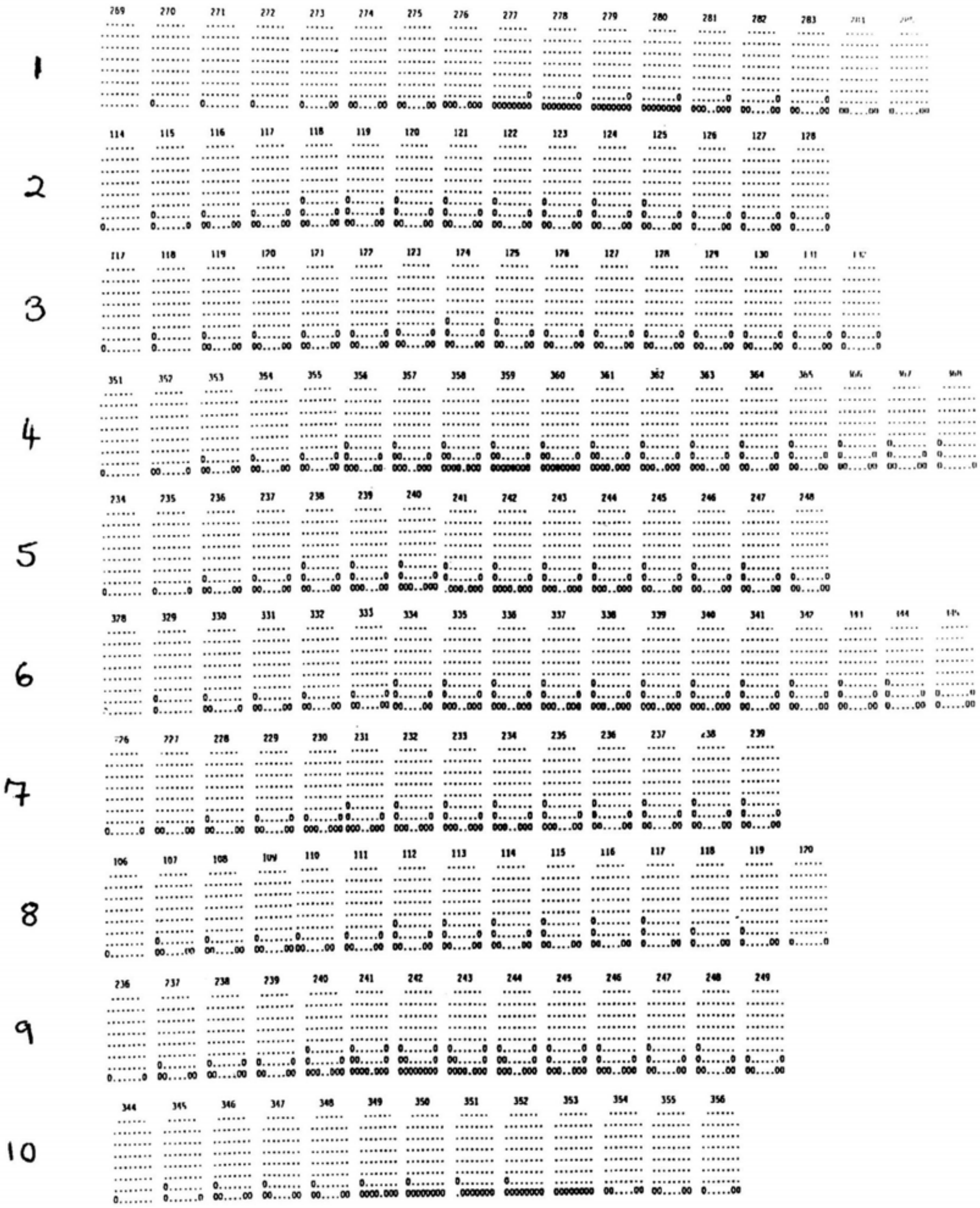


*Figure 3.4 (viii): **subject H** EPG patterns for 10 fast speech /n#k/ repetitions*

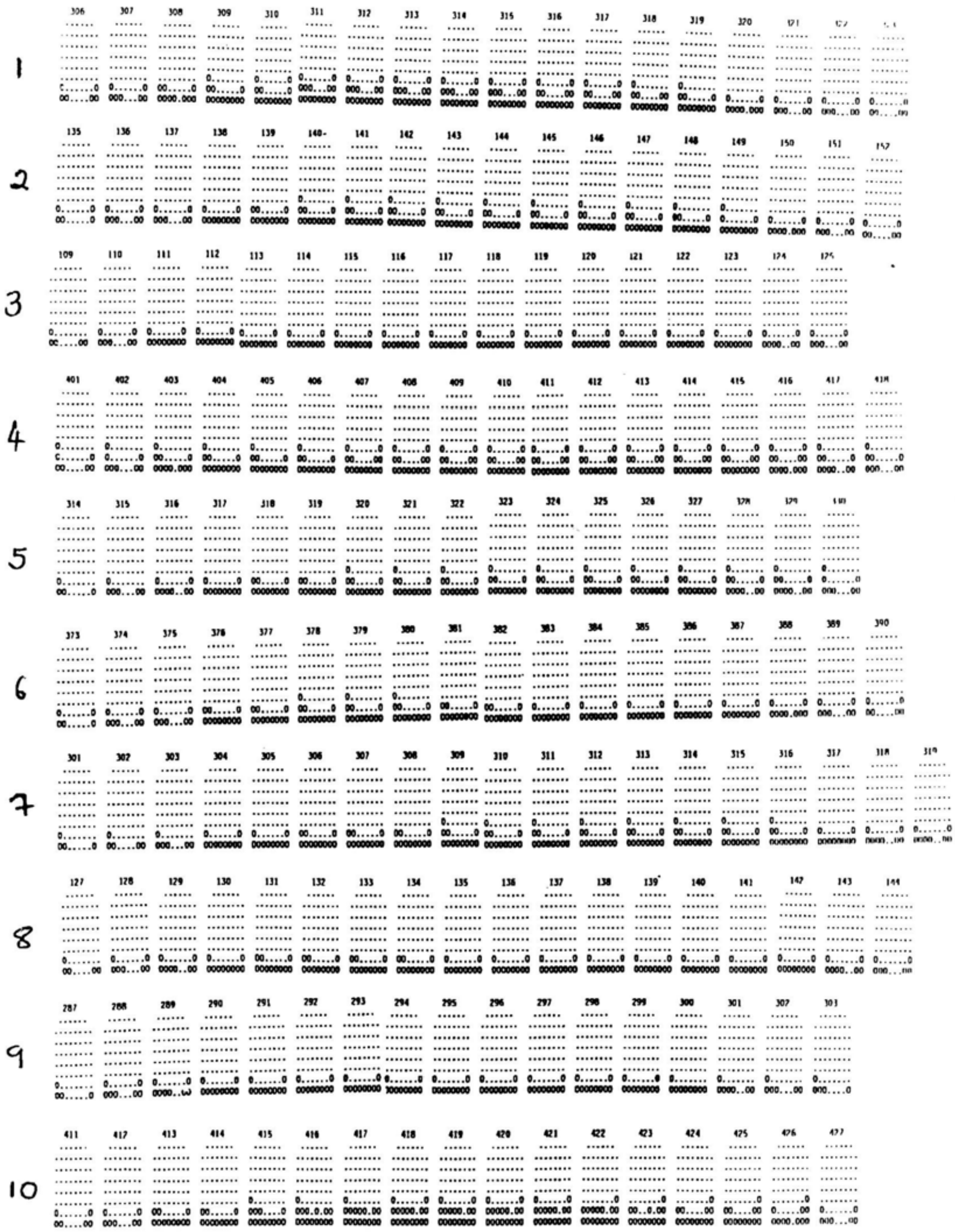


Figure 3.4 (ix): ***subject I*** *EPG patterns for 10 fast speech /n#k/ repetitions*

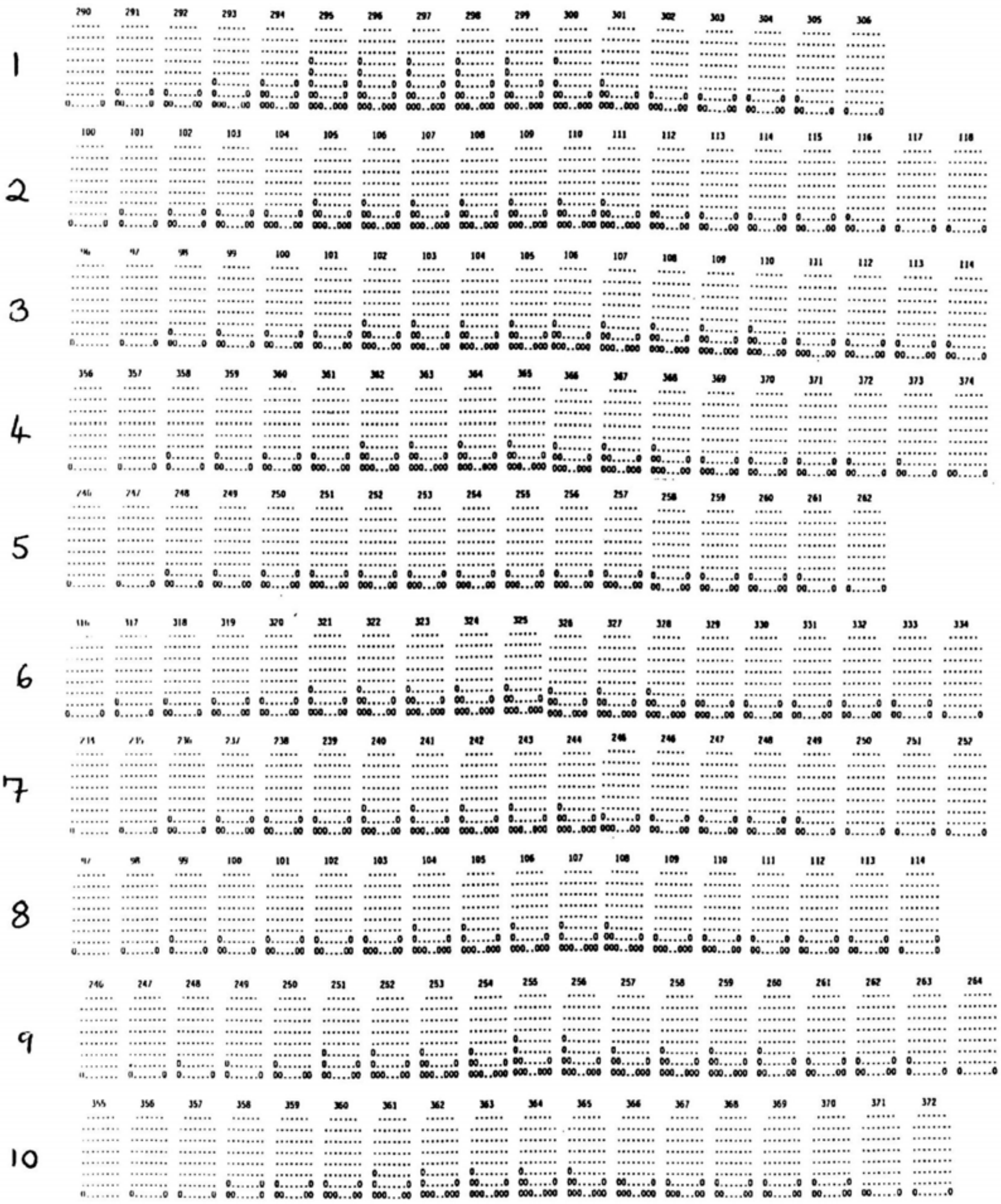


*Figure 3.4 (x): **subject J** EPG patterns for 10 fast speech /n#k/ repetitions*

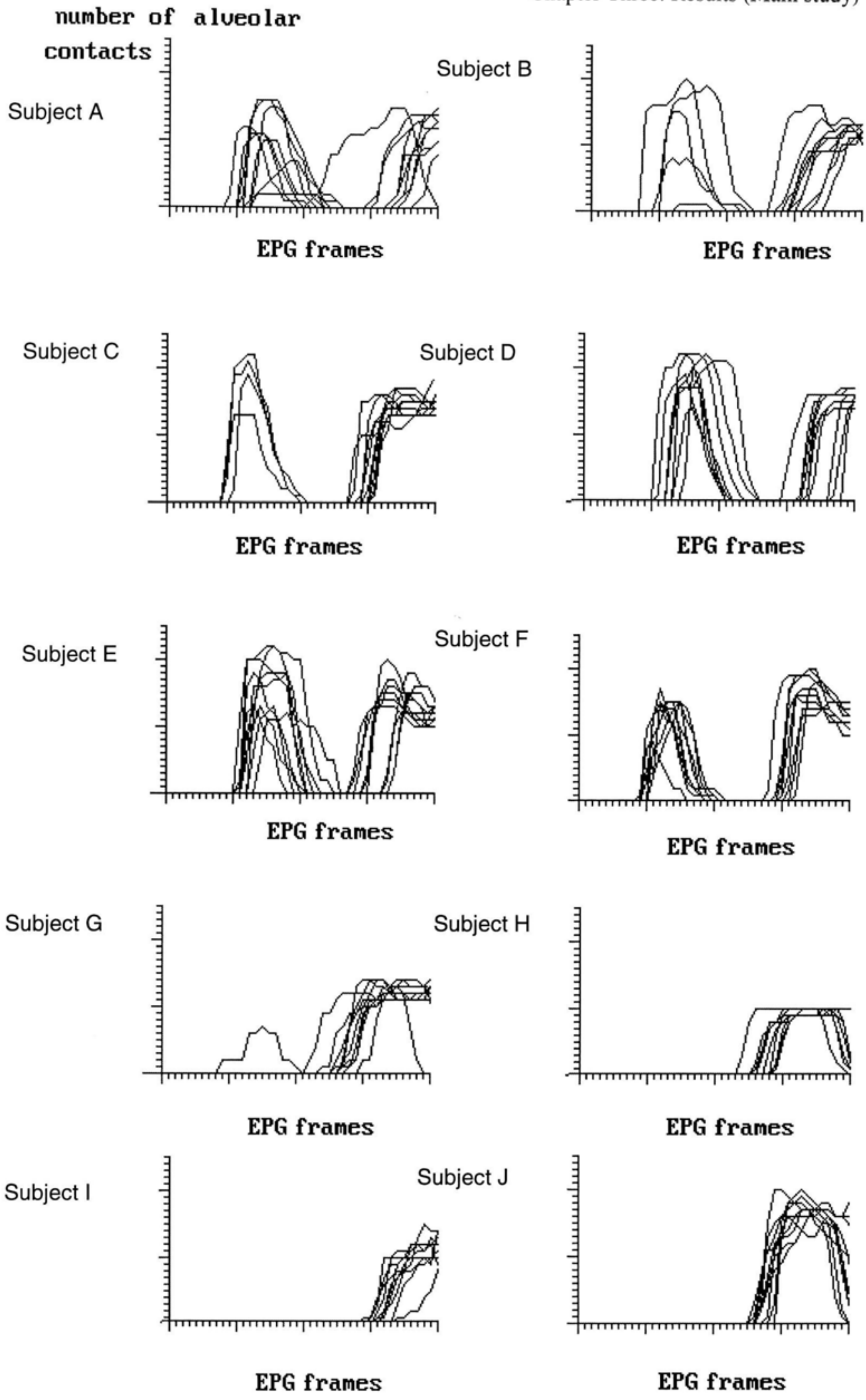


Figure 3.5: 'contact totals' displays for subjects A-J. Each display shows number of electrodes contacted in the alveolar region (first 3 rows) during /an#kʌts/ fast speech - All 10 amount-of-contact curves for each subject are superimposed onto a single display - Time is represented in the x-axis, one frame=10 ms.

3.1.2.1 Subjects G, H, I and J

Subjects G, H, I and J assimilated all 10 repetitions of fast speech /n#k/. Apart from one repetition produced by subject G, the EPG patterns for these subjects, Figure 3.4 (vii)-(x) show a complete lack of an alveolar gesture suggesting that an habitual fast speech phoneme substitution process is at work. There appears to be no influence of the alveolar gesture on the velar gesture even in the form of a residual tongue body movement. This residual movement would be manifested in the EPG data as side tongue or 'lateral' contact on the EPG display stretching further forward than contact characterising a lexical velar articulation. The assimilated alveolar to velar sequences appear to be indistinguishable from the lexical velar to velar *...bang comes...* tokens. This group of 4 speakers shall from here on be referred to as '100% categorical assimilators' ('100%' refers to the fact that all repetitions were assimilated)

3.1.2.1.1. Differences between underlying and derived /ŋ#k/

This section reports on EPG measures (spatial and timing) and acoustic measures which were taken to explore the possible identity of derived /ŋ/, i.e. complete assimilation with underlying coronal specification, and lexical /ŋ/, i.e. a specified velar.

In order to summarise and compare this group's place of articulation for lexical and derived /ŋ/, Figure 3.6 (i) & (ii) show a single schematic frame ('composite EPG displays') for each speaker representing the frequency and location of contacted electrodes at the *end of voicing* over 10 repetitions for each sequence (also refer to section 2.1.5.2.1 in Chapter Two). Since the frame of maximum velar closure was not annotated for any of the data, 'end of voicing' was considered an adequate representation of place of articulation for /ŋ/. Each type of sequence ('derived' /ŋ/ and lexical /ŋ/) is an all-velar sequence inolving a velar nasal followed by a voiceless velar plosive and so it was not anticipated that the place of articulation at the transition from the nasal to the plosive would vary much from the place of articulation earlier in the velar nasal. Frequency of electrode contact is indicated by the shading of the individual electrodes and by the actual number assigned to each electrode which corresponds to the number of times it was contacted over 10 tokens.

Figure 3.6 (i) - composite EPG frames: tongue contact at end of voicing ...ban cuts...fast speech

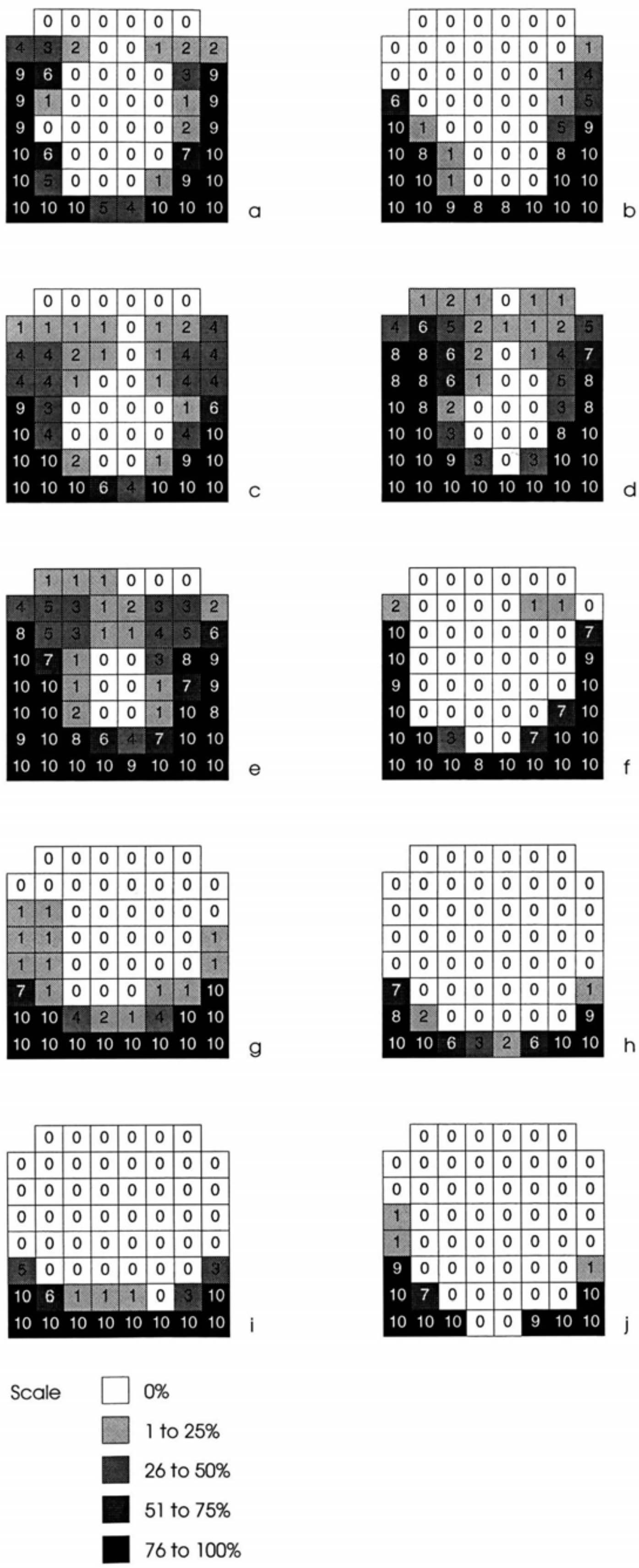
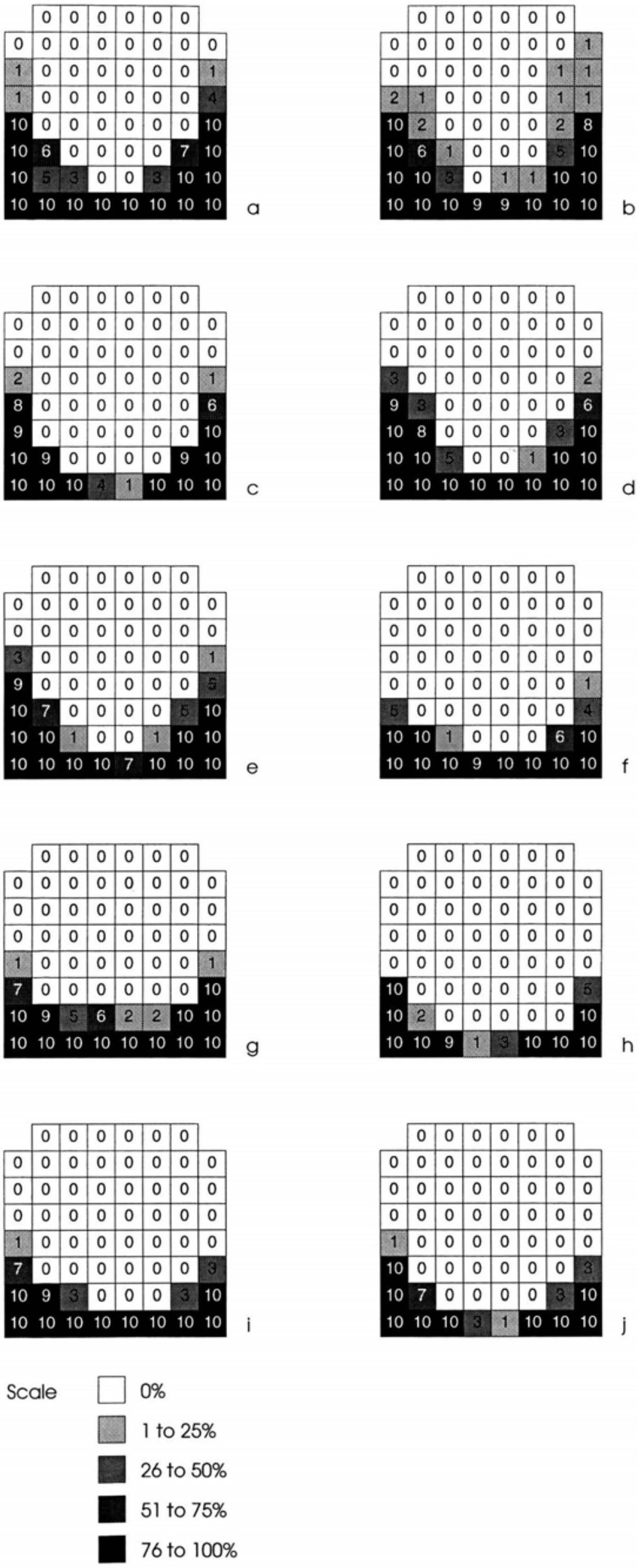


Figure 3.6(ii)-composite EPG frames: tongue contact at end of voicing...bang comes...fast speech



The /n#k/ composite EPG displays for the 100% categorical assimilators (G, H, I and J) appear to be very similar to those for /ŋ#k/ indicating a strategy of discrete assimilation. The contact totals displays in Figure 3.5 for these subjects summarise the absence of the alveolar gesture for fast speech /n#k/. There are however some differences in the composite displays which require explanation. For subject G, the display for /n#k/ (Figure 3.6 (i)) shows a number of electrodes contacted once as far forward as row 3 into the alveolar region. This at first suggests a single production of a residual alveolar but, as mentioned in section 3.0 above, for one token the alveolar and velar sequencing was anomalous. This can be seen in the EPG patterns, Figure 3.4 (vii) for this subject, repetition 9. Some build-up for an alveolar closure starts at frame 418 during which a velar gesture is produced at frame 421. 10 ms after the velar closure is formed alveolar closure is fully made. Since the alveolar closure and release occurs during the velar closure, the percept is of an assimilation. Once this is taken into account the composite place of articulation displays for subject G's fast speech /n#k/ and /ŋ#k/s are almost identical. Subjects H, I and J also have very similar displays, although a note must be made concerning subject J. For /n#k/ one repetition involves lateral contact up into row 4, which is one row in advance of the farthest forward contacted row for the control /ŋ#k/ display. This would at first sight suggest that one /n#k/ token satisfies the criteria set down for the data analysis for a residual alveolar articulation (see section 2.1.5.1, Chapter Two). However, on closer inspection of the /ŋ#k/ EPG data for this subject, it turns out that repetition 4 shows side contact as far forward as row 4. This means that all /n#k/ tokens can be regarded as complete assimilations. Unusually, for the /ŋ#k/ repetition 4, the annotation point 'end of voicing' which the composite displays are based on has not captured the maximum constriction for this velar closure. Figure 3.7 shows the EPG screen display for this token. The EPG frame at the top shows the onset of maximum constriction for /ŋ/. Maximum constriction lasts for approximately 30ms and decreases before the offset of voicing (labelled as 'env' on waveform).

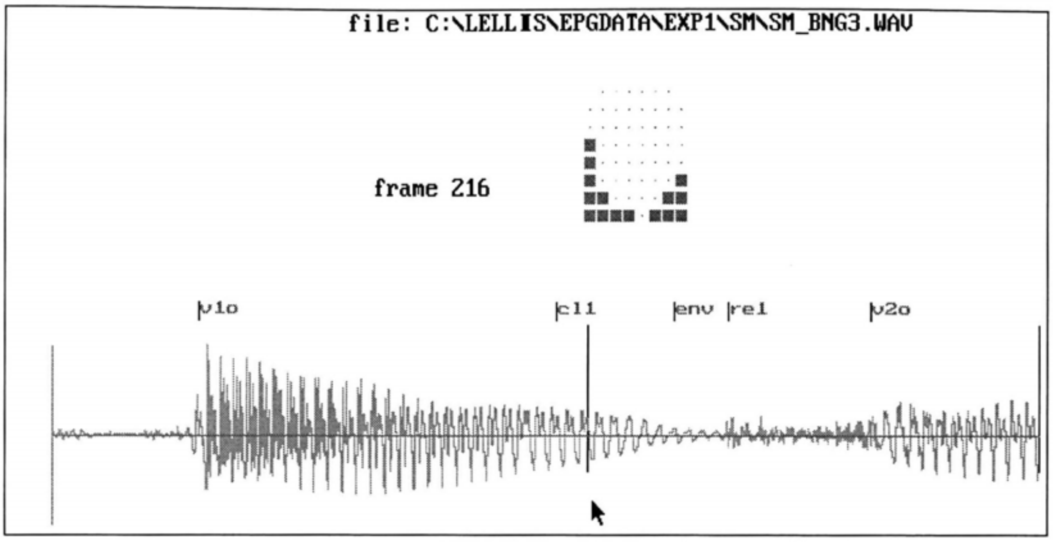
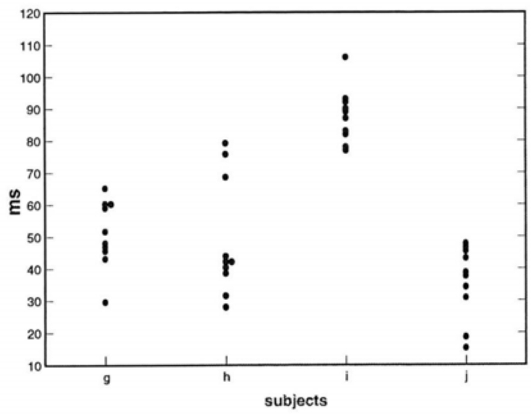


Figure 3.7: EPG screen display of repetition 4 of fast speech /ŋ#k/ produced by subject J. Onset and offset of maximum velar constriction (shown above waveform) occurs before end of voicing.

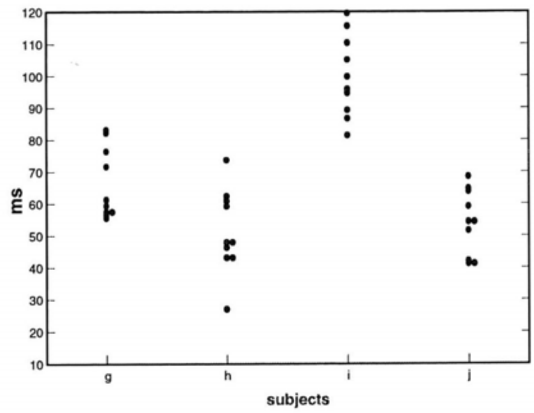
A final observation to be made from the composite EPG displays for derived and lexical /ŋ/ is that three of the four 100% categorical assimilators sometimes produce more retracted velar patterns for derived /ŋ/ than for lexical /ŋ/. For instance, for subject H /n#k/, side contact for 7 assimilated /n/ tokens reach row 6 on the left hand side of the display and only one token does on the right hand side whereas for the same subject, side contact for all productions of lexical /ŋ/ reach row 6 on the left hand side and the right hand side also shows more forward contact. This could be an indication that the tongue configuration for the retraced tokens is appropriate for a raised tongue tip that falls short of the alveolar ridge for an undershot alveolar closure, as Wright and Kerswill (1989) have hypothesised. They have speculated about a particular tongue configuration where: 'the tongue tip moves up towards the alveolar ridge, the blade and pre-dorsum become concave, which reduces the amount of lateral contact in the pre-velar area. At the same time, this tongue shape will cause the velar contact itself to be more retracted.'. With tongue-palate contact data only it is not possible to confirm or disprove this. Chapter Four presents more details about this issue.

The possibility that the radical assimilation strategy used by these four speakers can be attributed to a faster rate of speech can be assessed by considering Figure 3.2 above. The 10 fast speech values for each subject lie beneath the line. While subject H and J *are* speaking more quickly compared to all others, whose values stay mainly above the 400 ms mark, subject G and I's values are comparable with the others.

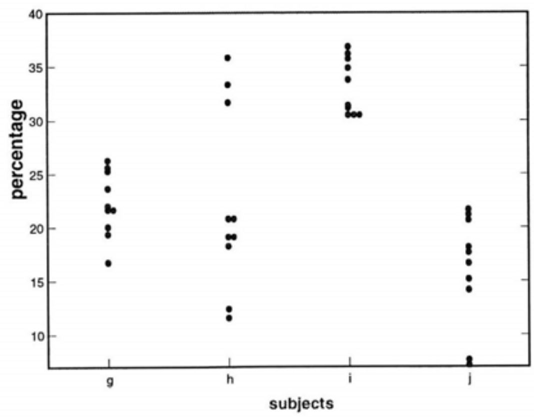
Another measure which might identify the two forms of /ŋ#k/ as distinct is the duration of voiced velar closure /ŋ/. If /n/ assimilation is complete, as the place of articulation data suggests, then it might be expected that duration of closure for the derived /ŋ/ is identical to that for lexical /ŋ/. Figure 3.8 compares the duration in ms of the velar closure for both forms measured as the interval between onset of velar closure (EPG-defined) and the end of voicing. Figure 3.8 (i) shows the durations of the assimilated /n#k/ velars and Figure 3.8 (ii) shows the same for the lexical /ŋ#k/ velars produced by the 100% categorical assimilators. Underneath, Figure 3.8 (iii) and (iv) show the same values for derived and lexical /ŋ/ respectively but this time expressed as a percentage of the duration between onset of the first vowel /a/ and the end of /s/ in ...*ban cuts*..., thus factoring out speech rate variations that might occur between and within speakers' tokens.



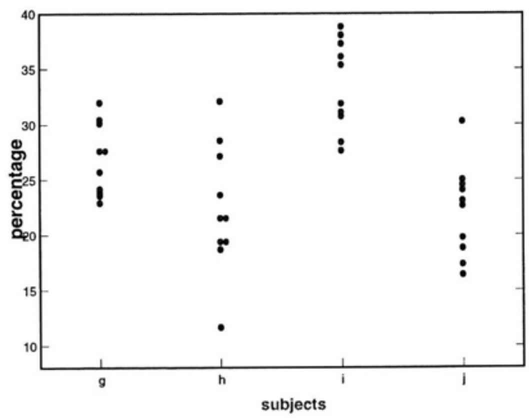
(i) assimilated /n#k/



(ii) lexical /ŋ#k/



(iii) assimilated /n#k/ *%*



(iv) lexical /ŋ#k/ *%*

Figure 3.8 (i) and (ii) = duration of velar closure in ms
(iii) and (iv) = % duration of velar closure

Table 3.6 shows the means and standard deviations for the percentage lexical and derived velar closure values. Pooled means are shown first followed by means for individual subjects.

Table 3.6 means and standard deviations for % duration of all lexical /ŋ/s ('/ŋ/') and all derived /ŋ/s ('/n/') produced by subjects G-H. Pooled data is shown uppermost and means for individual subjects is shown below

Form	Cases	Mean (%)	Standard Deviation
/ŋ/	40	26.4	6.2
/n/	40	23.5	8

subject × form	Cases	Mean (%)	Standard Deviation
G × /ŋ/	10	26.9	3.3
G × /n/	10	22.3	3
H × /ŋ/	10	22.7	5.9
H × /n/	10	22.5	8.4
I × /ŋ/	10	33.7	4.1
I × /n/	10	33.2	2.6
J × /ŋ/	10	22.4	4.2
J × /n/	10	16.2	5.2

Separate variance t-tests (because samples were known to have come from separate conditions) on the percentage values (Table 3.6) showed that duration of velar closure for assimilated /n/ and for lexical /ŋ/ was not statistically significant for subjects H and I (a Shapiro-Wilk test confirmed that all samples had a normal distribution). However, the difference between these two forms was highly significant for subjects G and J. The mean durations shown in Table 3.6 underline this effect.

Table 3.7 p values for derived and lexical /ŋ/ being equal, subjects G-J

subject	**G**	**H**	**I**	**J**
probability	0.0044 *SIG*	0.9504 *NOT SIG*	0.7778 *NOT SIG*	0.0093 *SIG*

Comparing *all* derived and lexical /ŋ/ percentage durations (fast speech), including those produced by subjects A, B, C and D but excluding residual alveolar articulations (those where there is no mid-sagittal contact for target stop closure), a t-test showed that these two forms of /ŋ/ are not significantly different. All samples had a normal distribution, confirmed by Kolmogorov-Smirnov test.

3.1.2.2 Subjects A, B, C and D.

It has been noted that subjects A, B, C and D sometimes assimilate and sometimes do not in fast speech. Figure 3.3 shows that subjects B and C assimilate the same number of times and so it might be inferred that they are applying the same assimilation strategy. However, this simple assimilation versus non-assimilation graph does not tell the whole story. The alveolar contact totals displays for these subjects in Figure 3.5 indicate for all 4 speakers the absence of alveolar contact for some repetitions, but the displays differ with respect to variability in amount of contact for those repetitions which register some contact in the alveolar ridge. The contact totals curves for subjects A and B are more variable in the y-axis than those for subjects C and D, an impression brought about by the presence of progressively shallower curves on the displays with some curves registering only one or two contacts in the alveolar region. Turning to the raw EPG displays for these speakers in Figure 3.4 (i)-(iv) it is clear that subjects A and B produce a spatial range of alveolar targets from full alveolar stop closure to complete assimilation, while subjects C and D do not. The latter group's patterns will be discussed first.

The assimilated tokens produced by subjects C and D are complete, that is, it appears that a phoneme substitution process has taken place as for the 100% categorical assimilators. Figure 3.6 (ii) shows subjects C and D's prototypical EPG frames for all 10 fast speech control /ŋ#k/ sequences. None of the complete assimilations in fast speech produced by subject C or D (Figure 3.4 (iii) and (iv) shows all EPG patterns) show contact on the palate farther forward than the place of articulation for any of the /ŋ#k/ tokens shown in Figure 3.6 (ii). Therefore, the /n#k/ assimilations can be considered complete assimilations and not residual alveolar gestures. The tokens where target /n#k/ is preserved shown in Figure 3.4 (iii) and (iv) show full alveolar stop patterns followed by velar closure. On the basis of these EPG data it would appear that subjects C and D have a binary opposition between full alveolar stop closure and complete assimilation. From here onwards these subjects will be referred to as the 'binary categorical assimilators'.

The complete assimilations are not a function of an increase in speech rate relative to the speech at which their full alveolar articulations are produced and so assimilation is in a real sense optional for subjects C and D. Figure 3.9 (i) below shows the speech rate in milliseconds for these subjects' non-assimilations (which appear on the graph as 'c non-assimilation' and 'd non-assimilation') and assimilations ('c assimilation' and 'd-assimilation'). Speech rate is measured in the same way as previously, the interval between the onset of the vowel /a/ in ...*ban cuts*...up to the end of /s/.

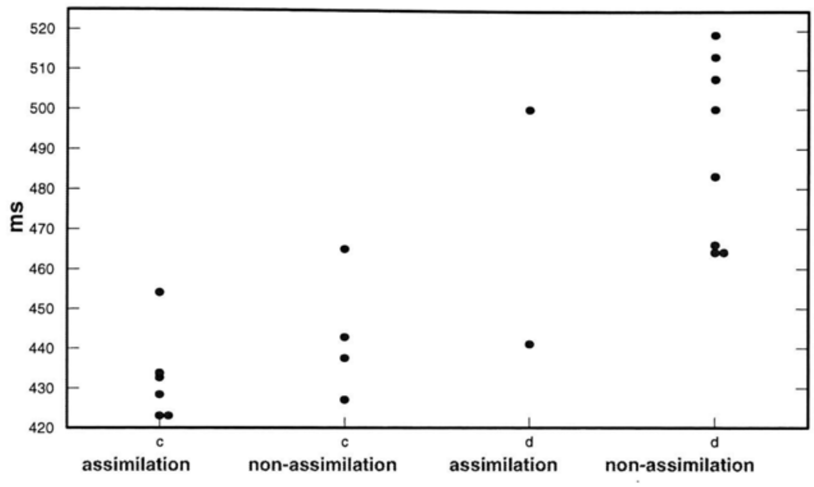


Figure 3.9 (i) speech rate (time in ms between /a/ in ...ban cuts...to end of /s/) of non-assimilations and assimilations produced by subjects C and D

This scatterplot shows that assimilations are only sometimes produced at a faster speech rate than non-assimilations and this applies to both subjects.

While the spatial variability in target alveolars for fast speech /n#k/ produced by the binary categorical assimilators form a strict binary opposition, that of subjects A and B forms a continuum including the all-important intermediate assimilatory forms. These forms correspond to the reduced contact totals curves in Figure 3.5. Figure 3.10 shows the EPG patterns for subject B ranked in descending order of alveolar contact to show a spatial continuum between full alveolar closure patterns to complete assimilations. In the same format as for the EPG pattern displays already referred to, there is one numbered repetition per line capturing roughly the onset of alveolar closure, or velar closure if an assimilation has taken place up to and including the release of the velar closure for /k/. Full alveolar closure patterns appear at the top of the page followed by patterns showing increasing amounts of assimilation and complete assimilation patterns appear at the bottom. The residual alveolar patterns between the two strategy extremes show either incomplete closure on the alveolar ridge with no mid-sagittal contact (number 4 on Figure 3.10) or show extensive lateral contact (number 5).

Frequency and location of contact at the end of voicing (indicating place of articulation for /ŋ/) for subject A and B's 10 repetitions of *...bang comes...*/ŋ#k/ are shown in the composite EPG frames, Figure 3.6 (ii). The frame for subject B shows that for one token side contact for /ŋ/ reached as far as row 2. This token (repetition 5 of fast speech /ŋ#k/) is shown below.

Figure 3.9 (ii) Subject B repetition 5 fast speech /ŋ#k/

This pattern shows an unusually 'forward' lateral extension and especially because there is some mid-sagittal contact normally associated with a coronal target at frame 278-279, the token was discounted as a 'canonical' /ŋ/ and considered instead as a mis-articulation. With this token put to one side, maximum lateral contact for any fast speech /ŋ#k/ token produced by subject B extends up to row 4 (see left-hand side of palate Figure 3.6 (ii)). Therefore any of subject B's assimilated /n#k/ tokens with lateral contact farther forward than row 4 before the velar closure can be considered a residual alveolar. It is interesting to note here that if this /ŋ#k/ token was not discounted then it would appear that subjects A and B have the most spatial variability for lexical /ŋ/ place of articulation compared to all other subjects (Figure 3.6 (ii)). This would be of potential importance to this study. This would suggest that /n#k/ residual alveolars reflect a pervasively high level of speech production variability, as observed in subjects' A and B lexical /ŋ/ place of articulation summaries. Turning now to the fast speech lexical /ŋ/ place of articulation summary for subject A in Figure 3.6 (ii), we can see that lateral tongue contact on one occasion reached as far as row 3. This means that any /n#k/ token further forward than row 3 can be considered a residual alveolar.

Now that the normal spatial limits of lexical velars have been established, we are in a position to identify the occurrence of residual alveolars for subjects A and B. Repetition 5 on Figure 3.4 (i) of fast speech /n#k/ produced by subject A, at first sight appears to be a residual alveolar, that is, it has contact laterally extending into the alveolar region without closure on the alveolar ridge. However, since we have seen that lateral extension of control /ŋ#k/ for repetition 1 reaches row 3 for this subject we have to discount repetition 5 of /n#k/ as a residual alveolar and consider it a complete assimilation. Repetition 8 is, however, a residual alveolar. There is lateral contact on the left-side of the palate extending up into row 1, although there is insufficient mid-sagittal contact, according to the criteria, to constitute a stop occlusion. For subject B, we have seen how one /ŋ#k/ token is discounted which means that the outer limits for lateral contact is row 4. Thus /n#k/ repetition 6 (Figure 3.4 (ii)) produced by subject B can be identified as a residual alveolar. Repetition 7 lacks full closure on the alveolar ridge although is considered a non-assimilation on account of some mid-sagittal contact. This can be referred to as a token showing partial alveolar closure. In the entire database (all subjects) there are only 2 residual alveolar articulations which is something of a surprising finding. Subjects A and B will from here on be referred to as 'gradual assimilators'. In the light of the identification of these intermediate assimilation types a revised table of /n#k/ forms is shown in Table 3.8:

Table 3.8: occurrence of non-assimilations, residual alveolar articulations and complete assimilations in careful and fast speech for all subjects.

	careful		fast		
subject	**non-assimilation**	**assimilation**	**non-assimilation**	**residual alveolar**	**assimilation**
A	10	-	7	1	2
B	10	-	4	1	5
C	10	-	4	-	6
D	10	-	8	-	2
E	10	-	10	-	-
F	10	-	10	-	-
G	9	1	-	-	10
H	7	3	-	-	10
I	10	-	-	-	10
J	10	-	-	-	10

It is important to note that there could be a slightly different classificatory criterion for alveolar stops to the one used in this study. For this study, tokens showing incomplete closure on the alveolar ridge but with contact in the mid-sagittal area, see section 2.1.5.1, are considered to be alveolar stops and not residual alveolars. However, if the approach is taken that an alveolar realisation must show *complete* closure across the alveolar ridge in order to qualify as an alveolar stop (i.e. continuous left to right contacted electrodes in one or more rows in the alveolar region) then the subjects' output can be viewed slightly differently. This alternative approach would not affect the level of variability observed in fast speech for subjects A and B nor would it affect subject C who produced non-assimilations with ' complete' closure anyway. The alternative classification of subject D, however, is more involved. Repetitions 6 and 10 would be residuals due to the lack of activation for one electrode each, and hence, this subject no longer assimilates in a 'categorical' fashion in the same way as before (but would show gradience at the non-assimilating end of the binary option). A similar situation arises for subjects E and F. The overall results of this study – speakers have fundamentally different assimilation strategies – would be the same regardless of these different classifications. One of the primary reasons for adopting the 'mad sagittal' categoriser is so that incomplete closure due to the presence of an EMA tongue tip coil in the follow-up study is not mistaken for anything other than an alveolar stop and hence that there is comparability between each of the experiments described in this dissertation.

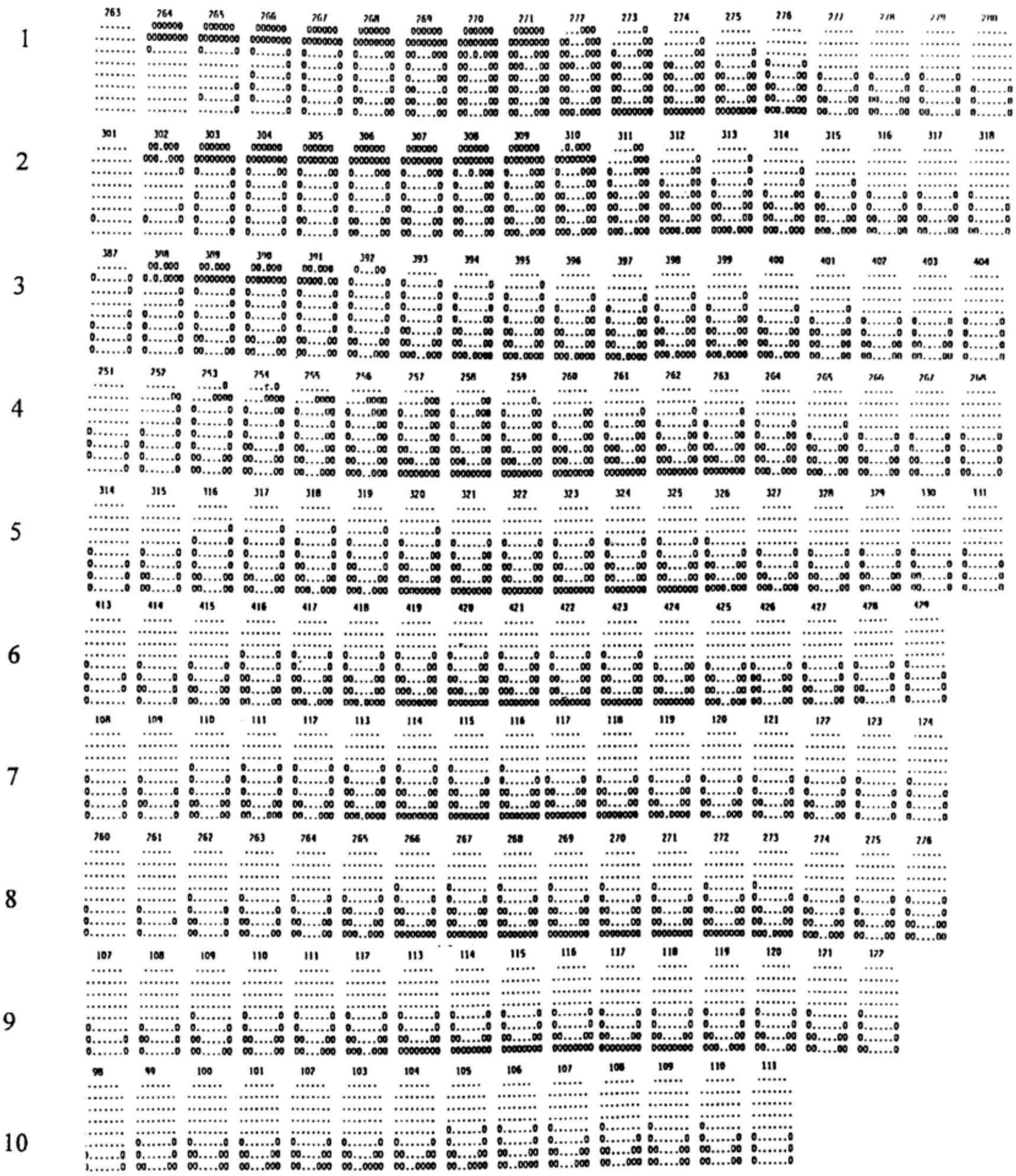


Figure 3.10 all 10 fast speech /n#k/ repetitions produced by subject B ranged in order of achievement of alveolar closure. Repetitions at the top are full alveolars and those at the bottom show maximal reduction of alveolar i.e. complete absence

At this point we must ask whether there is a relationship between articulation type (non-assimilation, complete assimilation, residual alveolar) and speech rate. As a corollary of this we must also ask whether subjects A and B, who produce all three articulatory types in fast speech have a greater spread of speech rate values compared to all other subjects who produce either two types or only one also in fast speech. Figure 3.11 below shows the speech rate of fast speech /n#k/ non-assimilations (unfilled circles), complete assimilations (red circles) and residual alveolars (green crosses) for all subjects. In answer to the first question, a separate variance t-test showed that the difference between the speech rate at which the fast speech complete assimilations were produced overall and the speech rate at which the fast speech non-assimilations were produced overall was highly significant (<0.001). A Kolmogorov-Smirnov test of normality showed that all samples had a normal distribution. However Figure 3.11 shows that this relationship is not altogether straightforward. There are some important interactions between *speaker* and articulation type. Subjects E and I appear to have a similar spread of speech rate values and yet one exclusively applied a complete assimilation strategy and the other exclusively applied a non-assimilation strategy. A separate variance t-test (the sample had a normal distribution) confirmed that there is in fact no statistical difference between these data sets.

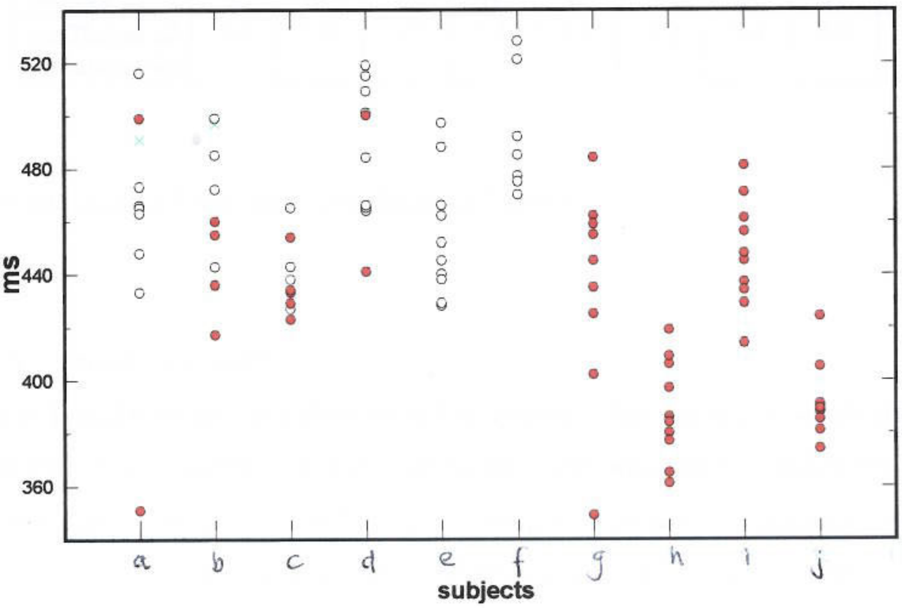


*Figure 3.11: scatterplot showing speech rate (measured as duration from onset of /a/ to end of /s/) of …*ban cuts*… fast speech for non-assimilations (unfilled circles), complete assimilations (red circles) and residual alveolar articulations (green crosses), for all subjects.*

The other observation that can be made from Figure 3.11 is that residual alveolars are not produced at faster speech rates since the two that were produced (subjects A and B) were produced in relatively slow speech. Furthermore none of the non-assimilated tokens produced by subjects A-F are produced below 420ms which appears to be something of a threshold.

It is clear from Figure 3.11 that the spread of speech rate for Subjects A and B, who produce all three /n#k/ articulation types in fast speech, is not greater than the spread of speech rate for all other subjects who either produced only one articulation type (E, F, G, H, I and J) or two (C and D). This means that the production of qualitatively different /n#k/ types is not a function of increased variability in speech rate. A coefficient of variation calculation for individual subjects confirmed this, although there is a tendency towards greater variability for subjects A and B.:

$$\frac{\text{Standard deviation}}{\text{mean}} \times 100$$

The higher the score for individual subjects, the greater the variability in speech rate across 10 tokens of fast speech /n#k/. Outliers produced by subject A and subject G (both outliers under 360ms), were left out. Table 3.9 shows the scores.

Table 3.9 coefficient of variance scores for speech rate of fast /n#k/ productions, all subjects

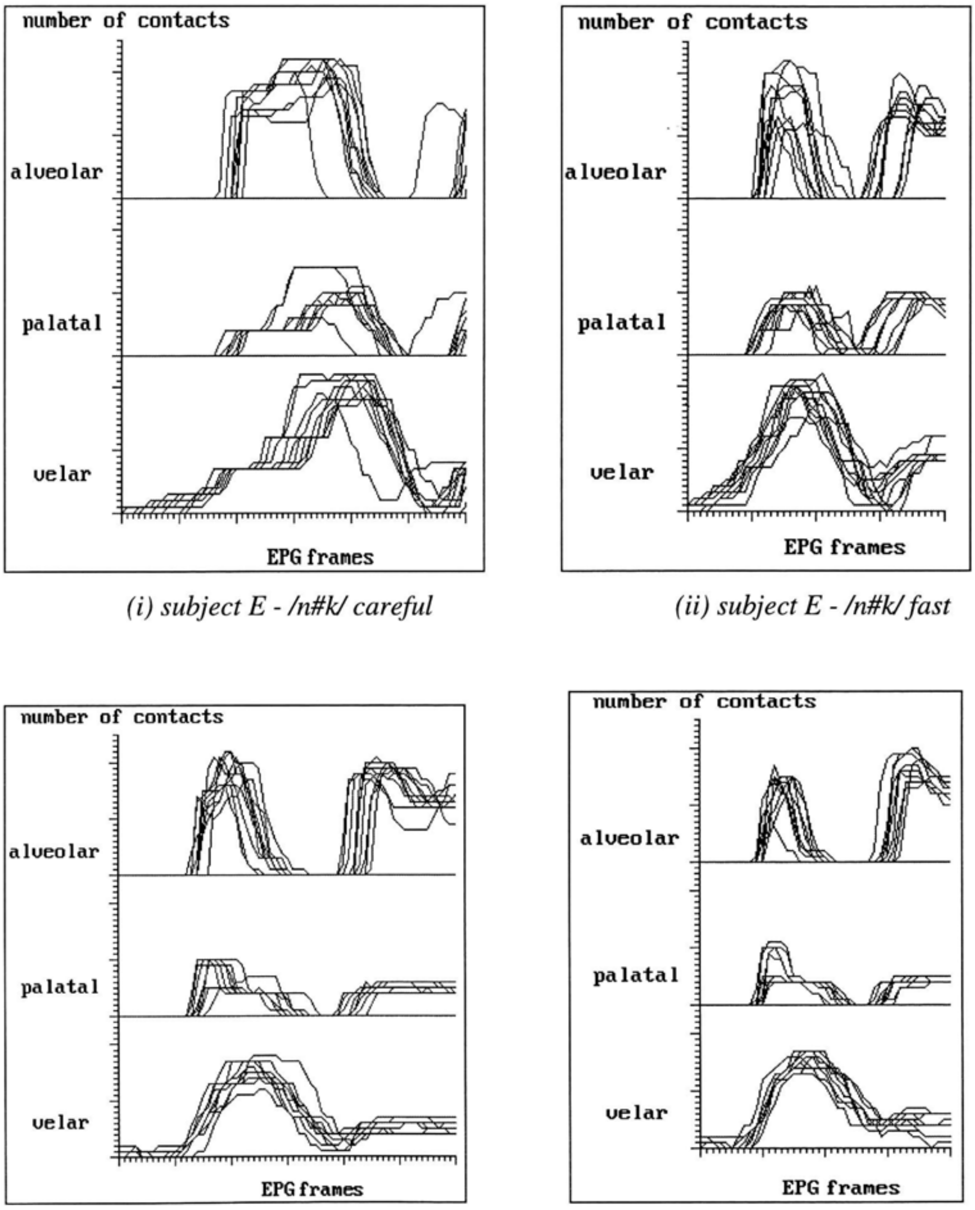
subject	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
coefficient of variance (%)	5.4	5.8	3.1	5.4	5.2	4.2	5.4	4.9	4.5	4.3

Means and standard deviations are shown in Table 3.2.

3.1.2.3 Subjects E and F

Subjects E and F do not assimilate /n/ in fast speech. Thus Table 3.1, which shows the overall occurrence of non-assimilation and assimilation, needs speaker-by-speaker breakdown to explain why only approximately 50% of fast speech tokens are assimilated across all speakers. Since these subjects do not use an assimilation strategy to cope with increased speech rate (or indeed for any other reason), it is necessary to compare their fast speech /n#k/ contact totals displays with those for careful speech /n#k/ to look into spatial changes in the production of this sequence. Figure 3.12 below shows contact in the alveolar, palatal and velar regions for /n#k/ in the two conditions (careful speech on the left) for subjects E and F.

The first thing to notice in the displays for the alveolar region is that subject E's alveolar stops (contact is registered on these displays from the onset of the vowel) in careful speech are roughly twice as long as the fast speech alveolars. The reduction strategy for this subject is to a large extent consonant shortening. In addition to stop closure shortening, there is increased variability in amount of contact across the alveolar region in fast speech with some repetitions showing roughly half the amount of contact than most careful speech tokens. The situation for subject F is different. It appears that there is less extensive shortening of the duration of the alveolar in the change from careful speech to fast speech, and less variability in amount of alveolar contact between fast speech repetitions.



(i) subject E - /n#k/ careful *(ii) subject E - /n#k/ fast*

(iii) subject F - /n#k/ careful *(iv) subject F - /n#k/ fast*

Figure 3.12 contact totals for subject E and F /n#k/ careful and fast speech

Another difference between these subjects seen from the contact totals in Figure 3.12, who shall be referred to as 'non-assimilators', lies in the time between initiation of the velar closing gesture relative to the initiation of the alveolar closure (otherwise known as temporal 'latency'). In careful speech subject E makes maximum velar closure, after a relatively long period of build-up, as the alveolar closure is released. In fast speech the velar curves slide leftwards on the x-axis with the result that maximum alveolar and velar contact appear to be synchronised. With subject F, however, maximum alveolar and velar contact appear to be synchronised at both rates of speech, with no adjustment from serial-ordering of maximum closure in each region for the slower rate to the eliding of these in fast speech.

Figure 3.11 shows that the 'non-assimilators' are not speaking at a slower rate of speech compared to other subjects. A possible explanation for the fact that subject F never assimilates in fast speech is that there is not sufficient increase in speed from careful speech to motivate a strategy switch. Figure 3.2 shows that subject F's careful and fast tokens are not clearly defined or separated out on the graph relative to other subjects. However, this explanation would clearly be wrong for subject E who has one of the widest speech rate discontinuities between the two speaking conditions and yet still does not take the opportunity to assimilate in fast speech.

3.1.3 Careful speech assimilations: EPG data

There were 4 assimilations produced in careful speech. Subject H produced three and subject G produced one. These subjects happen to be of the '100% categorical assimilators' group for fast speech. All four of these careful speech assimilations, like the fast speech assimilations, were complete, i.e. there was no evidence of the influence of an alveolar stop. The three assimilations produced by subject H are shown in Figure 3.13 below.

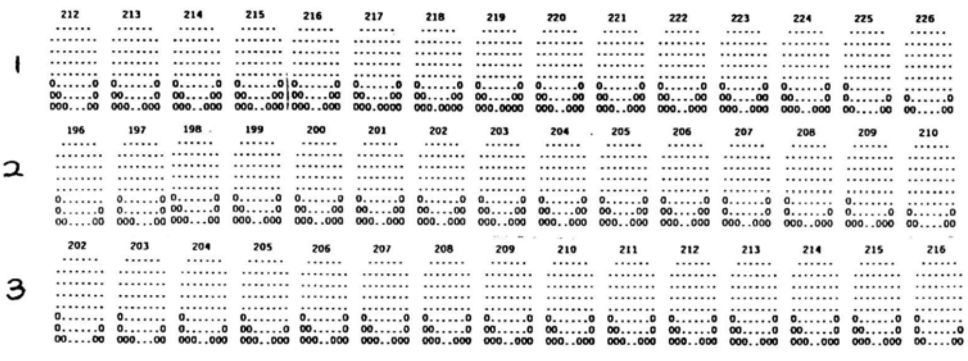


Figure 3.13: /n#k/ assimilations produced by subject H, careful speech

The one careful speech assimilation produced by subject G is shown in Figure 3.14 below. Prototypical EPG frames for these speakers' careful speech control velar-velar sequences are shown in Figure 3.15 below. For subject H none of the assimilated /n#k/ sequences show side contact extending further forward than row 6.

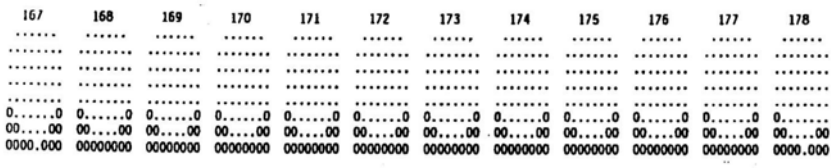


Figure 3.14: /n#k/ assimilation produced by subject G, careful speech

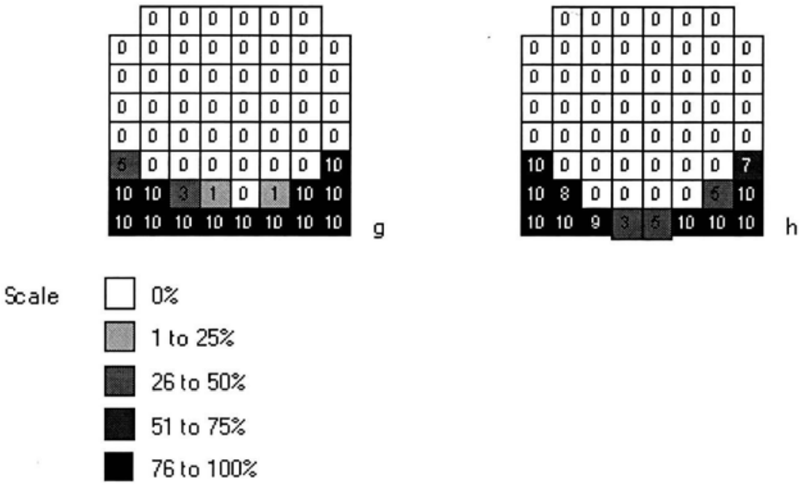


Figure 3.15 EPG prototypical frames for careful speech /ŋ#k/ at the end of voicing. subjects G and H

3.2 TIMING ASPECTS OF /n#k/

In this section, coordination of phonetic events for fast speech non-assimilated /n#k/, such as alveolar closure, velar closure, alveolar release, velar release, onset and offset of voicing is investigated. Indices of coordination include presence/absence of overlap of alveolar and velar stop closures and onset of velar closure following alveolar closure (referred to here as 'temporal latency'). The reason why these measures are important is that they provide a means of assessing (i) the intergestural timing involved in the production of alveolar to velar sequences, (ii) the effect that speech rate has on this intergestural timing, whether assimilation stems from changes in intergestural timing, and thus has a timing basis.

A full set of timing bars for all speakers can be seen in Figures 3.16 (i) –(x). The left-hand side of each page shows a single speaker's 10 careful speech repetitions (numbered) and the right hand side shows the same speaker's 10 fast speech repetitions plotted on the same scale. The 10 repetitions for each speaker are arranged in the order they were produced from 1 to 10 from the top of the page. Residual alveolars for subjects A and B are marked on the graphs and are represented by a velar bar only. This is because of the difficulty of measuring the onset and offset of these intermediate forms.

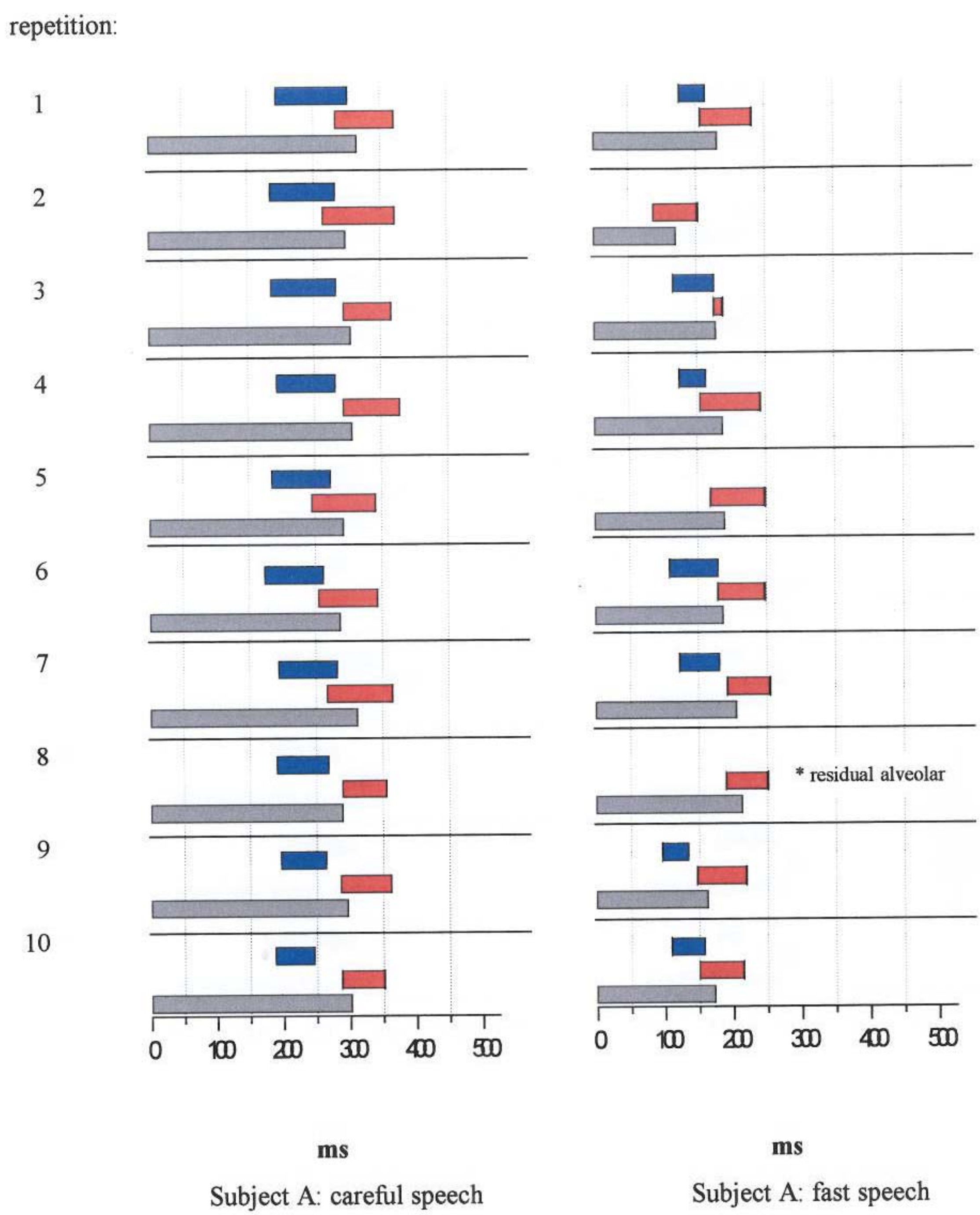
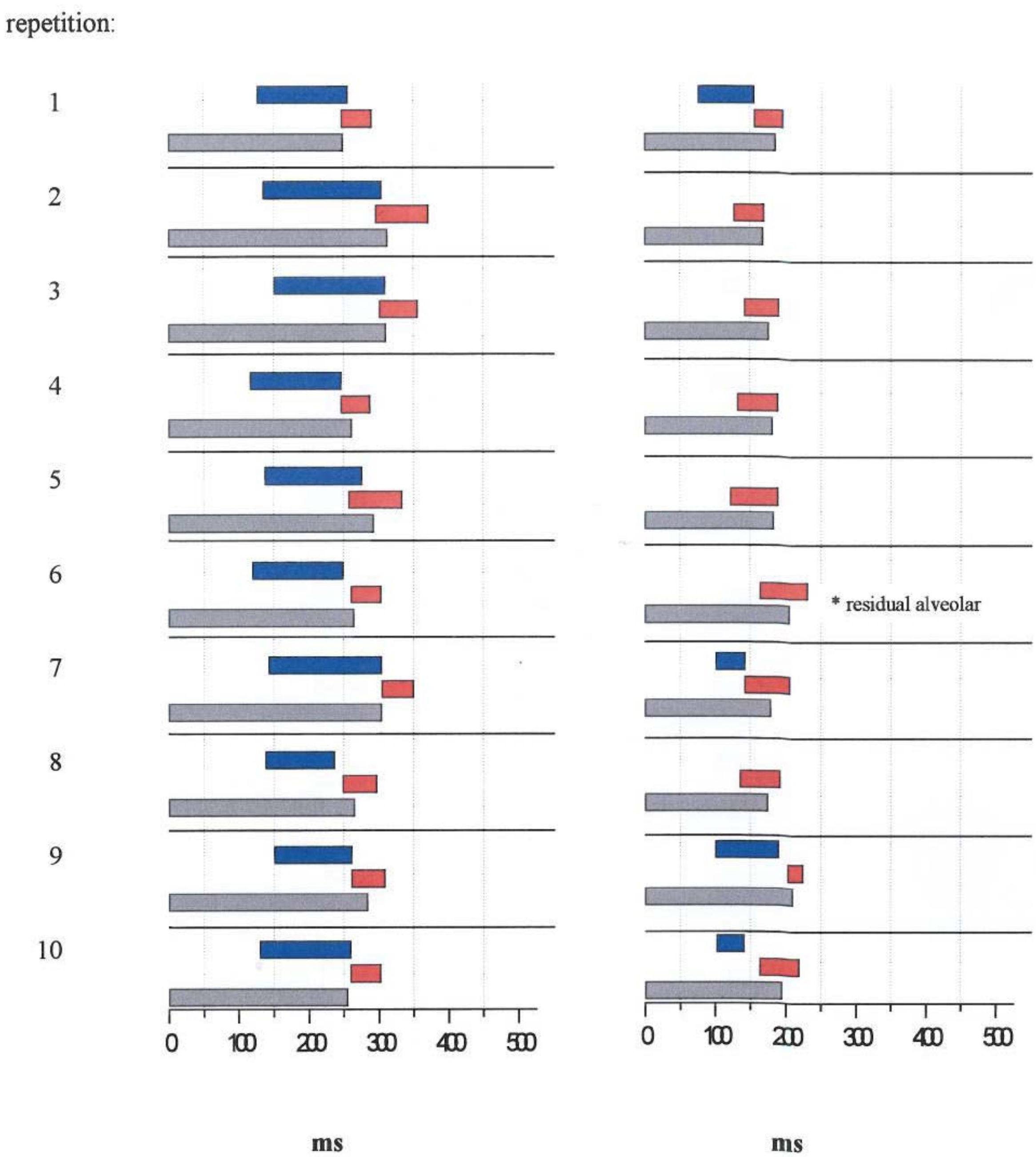


Figure 3.16 (i): timing bars showing coordination and duration in ms of phonetic events for all 10 repetitions of /n#k/ careful speech (left panel) and fast speech (right panel) from the onset of the vowel /a/ up to the release of the velar closure. Repetitions are numbered 1-10. Blue bars represent the alveolar stop (if produced), red bars represent the velar stop and the grey bar at the bottom represents voicing (starting at 0ms). The consonant bars are measured from EPG patterns and the voicing bar is measured from the waveform.



Subject B: careful speech

Subject B: fast speech

Figure 3.16 (ii)

(caption as for Fig. 3.16 (i) above)

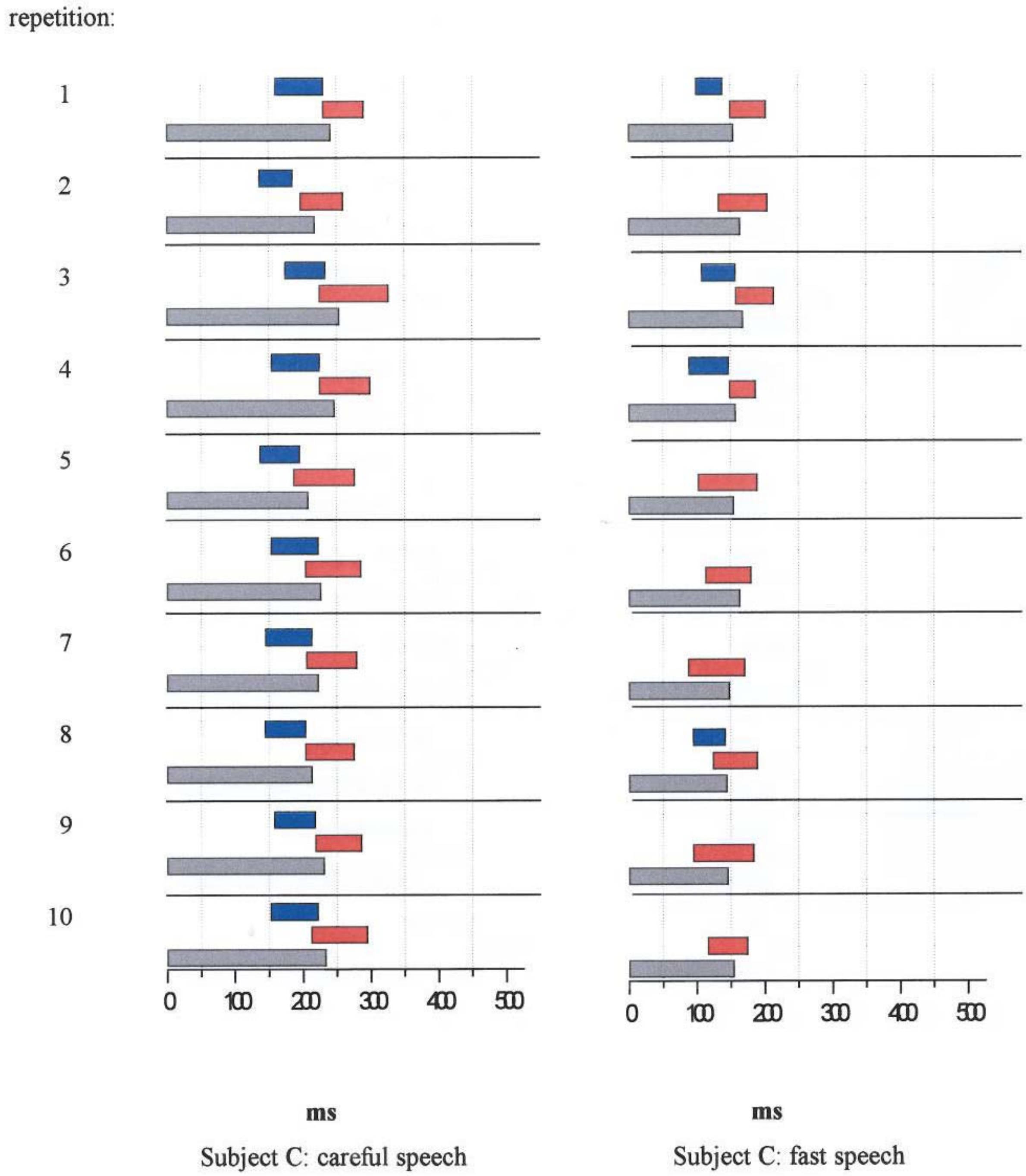


Figure 3.16 (iii)

(caption as for Fig. 3.16 (i) above)

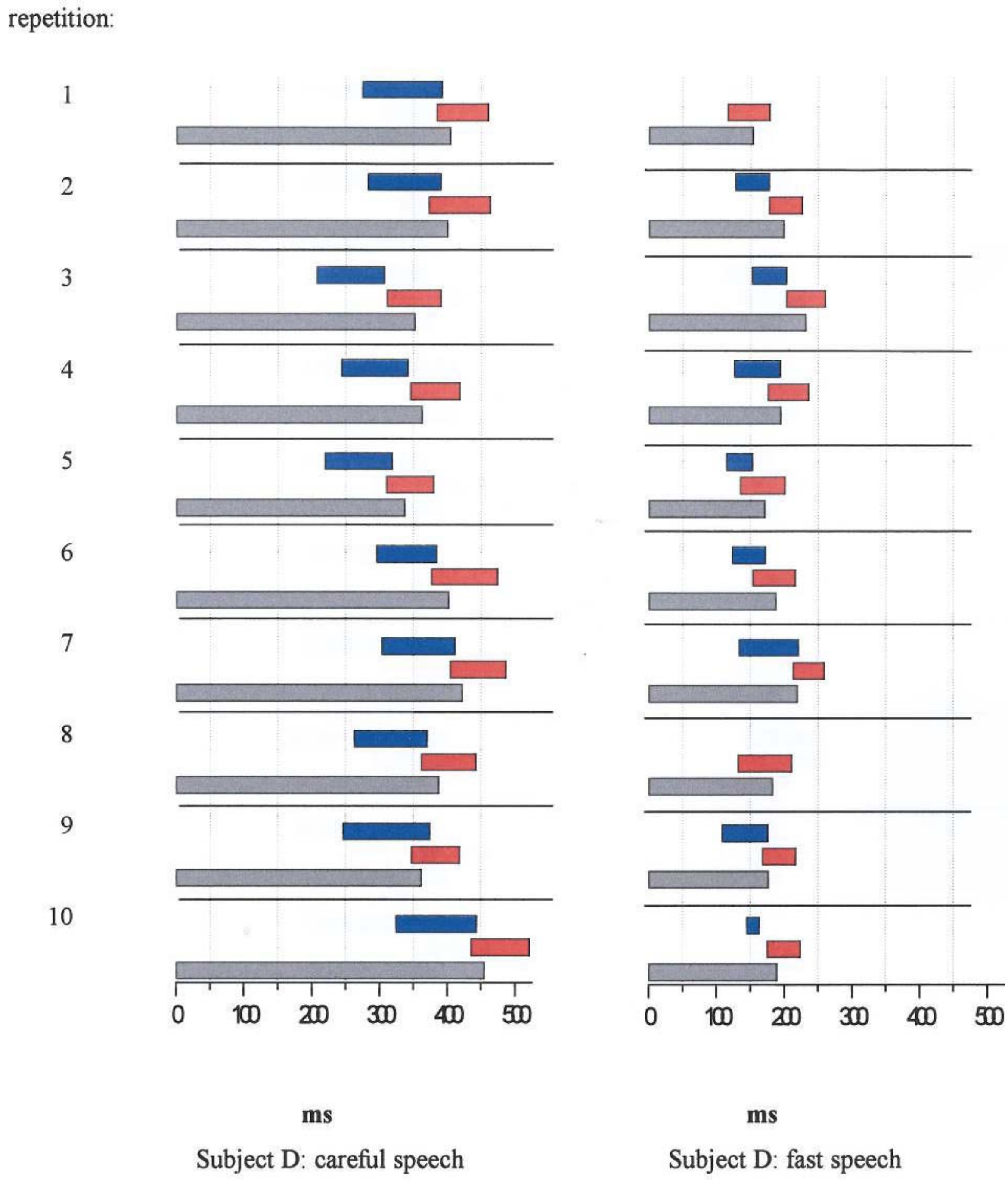
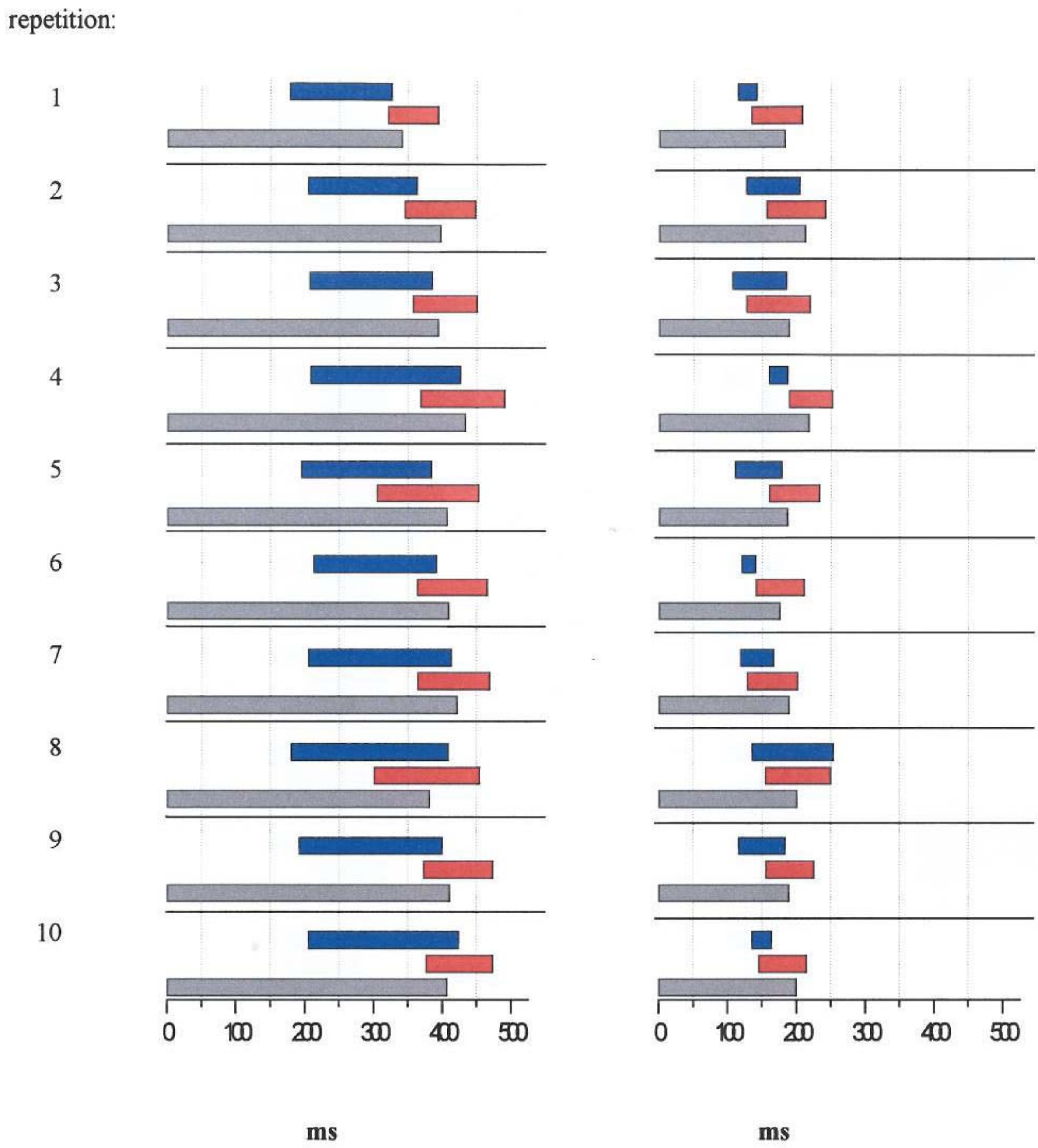


Figure 3.16 (iv)
(caption as for Fig. 3.16 (i) above)



Subject E: careful speech

Subject E: fast speech

Figure 3.16 (v)

(caption as for Fig. 3.16 (i) above)

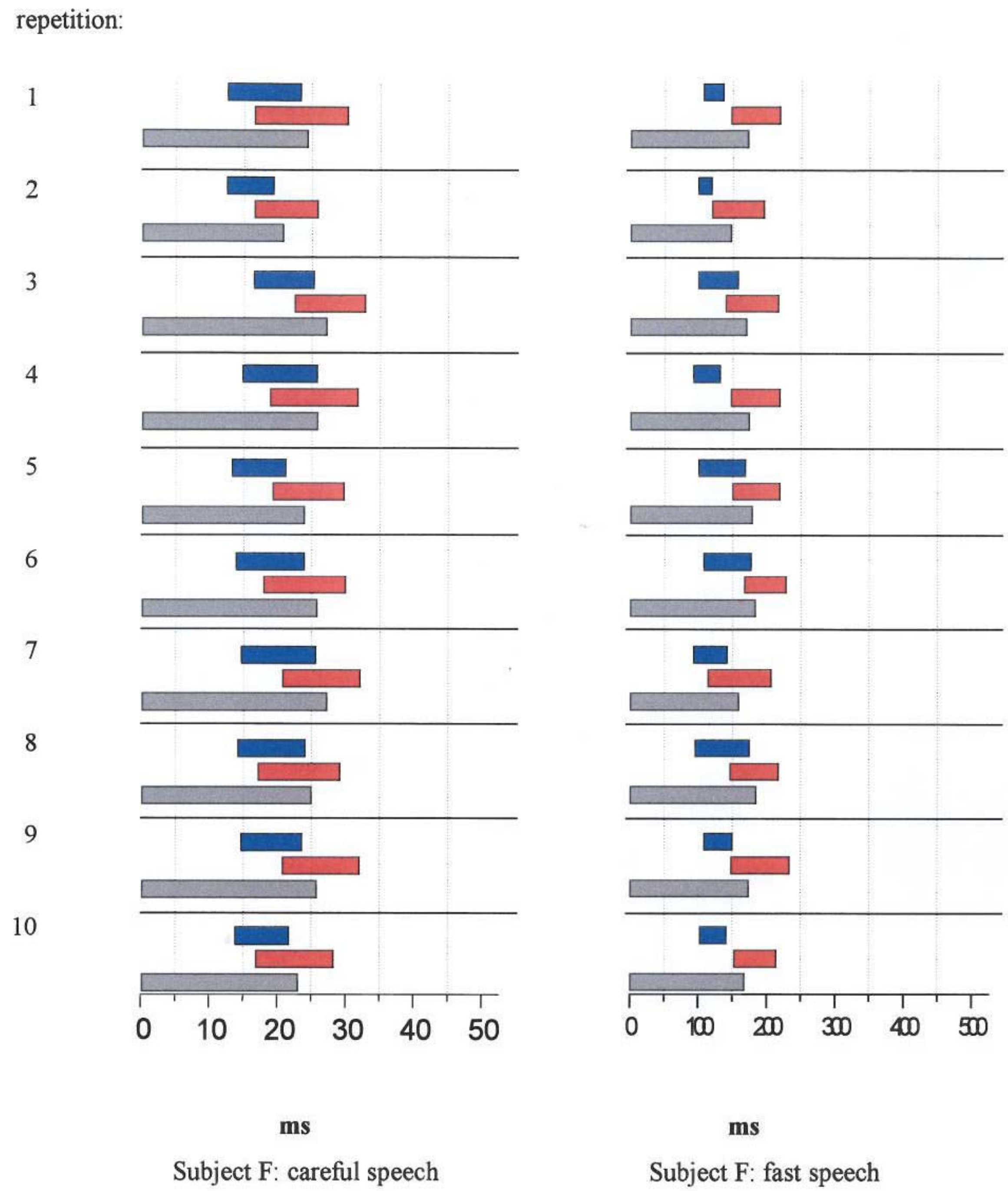
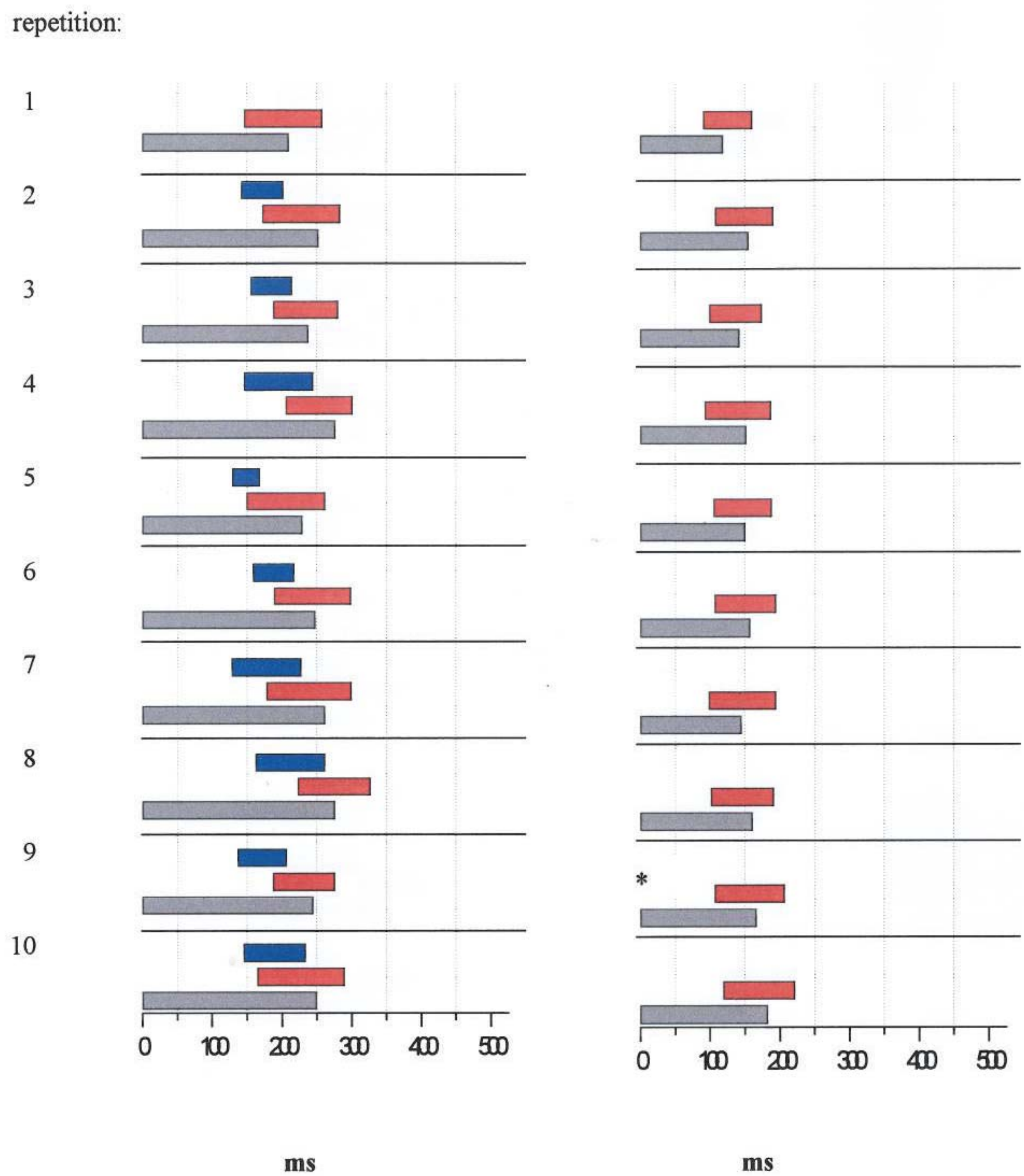


Figure 3.16 (vi)

(caption as for Fig. 3.16 (i) above)



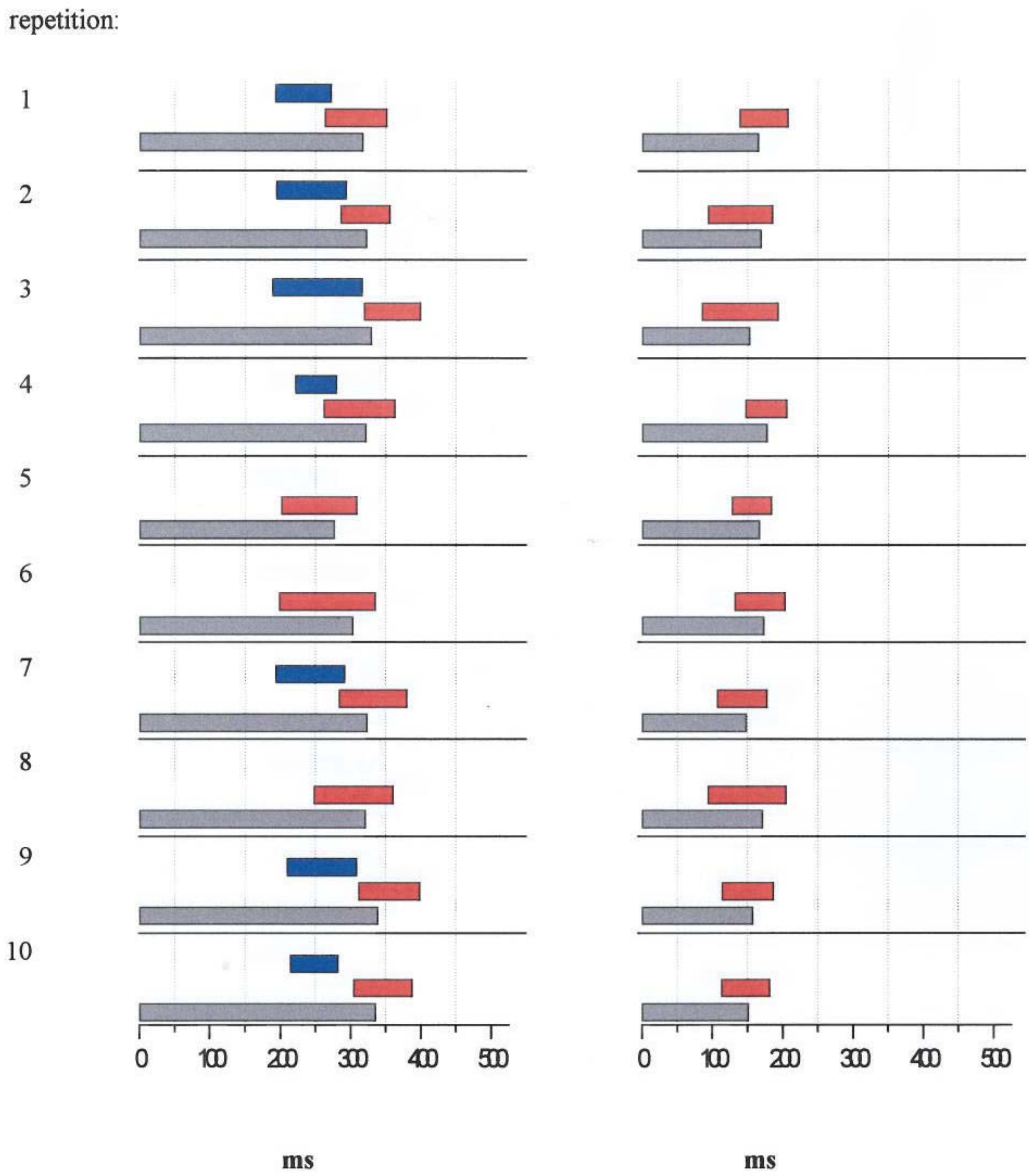
Subject G: careful speech

Subject G: fast speech

Figure 3.16 (vii)

(caption as for Fig. 3.16 (i) above)

** refers to anomalous production described in section 3.0, p.98*



Subject H: careful speech

Subject H: fast speech

Figure 3.16 (viii)

(caption as for Fig. 3.16 (i) above)

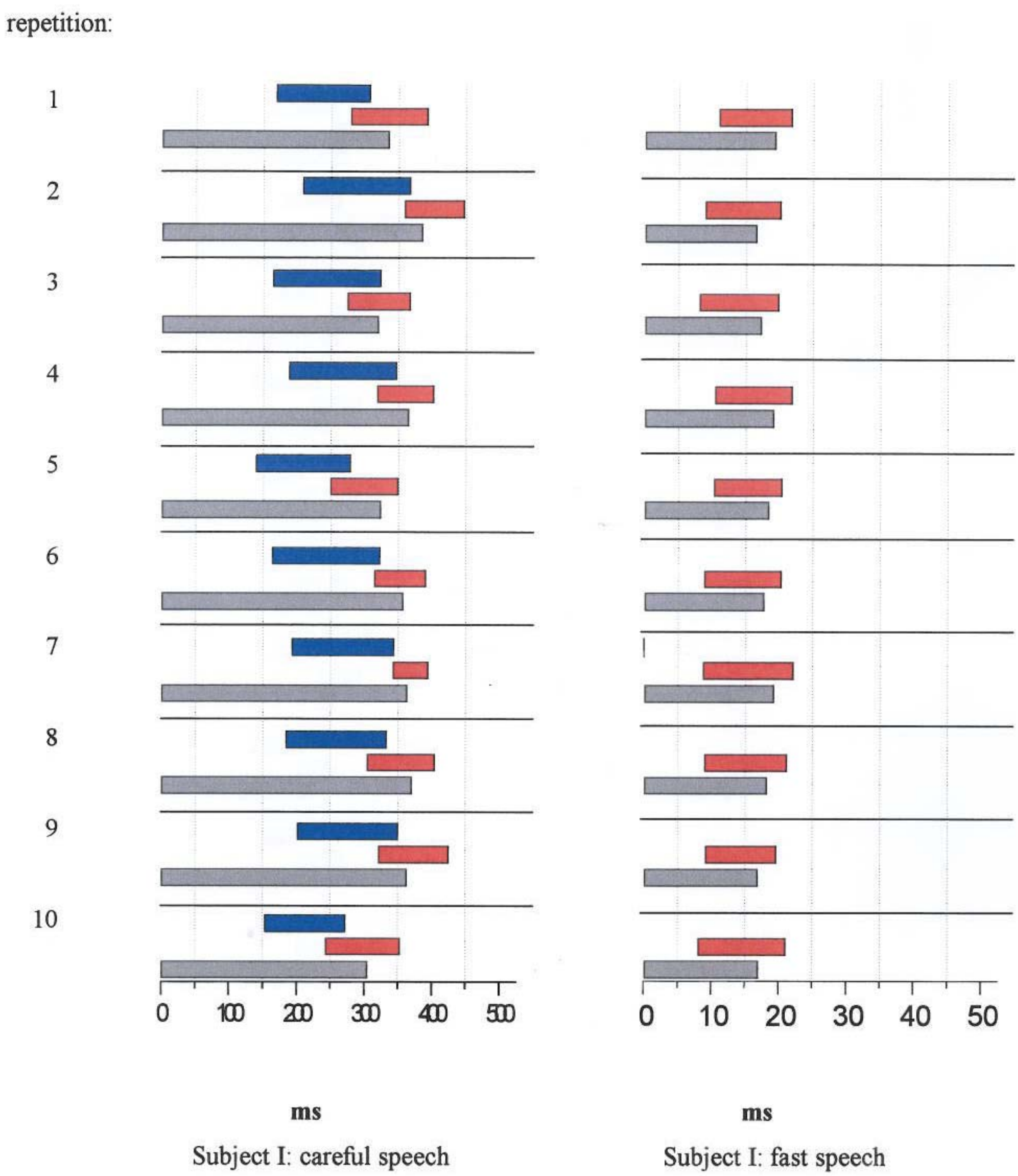
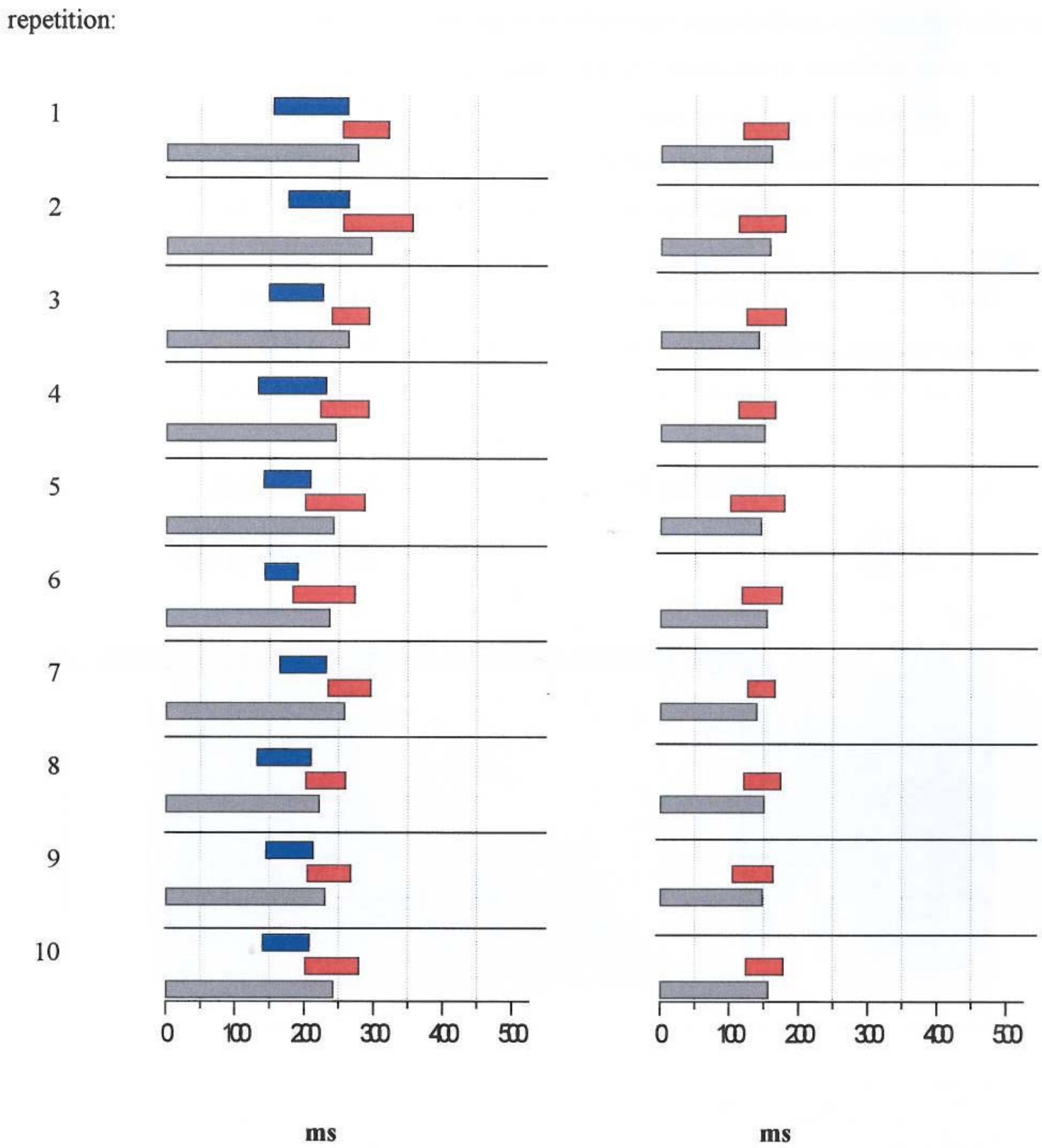


Figure 3.16 (ix)

(caption as for Fig. 3.16 (i) above)



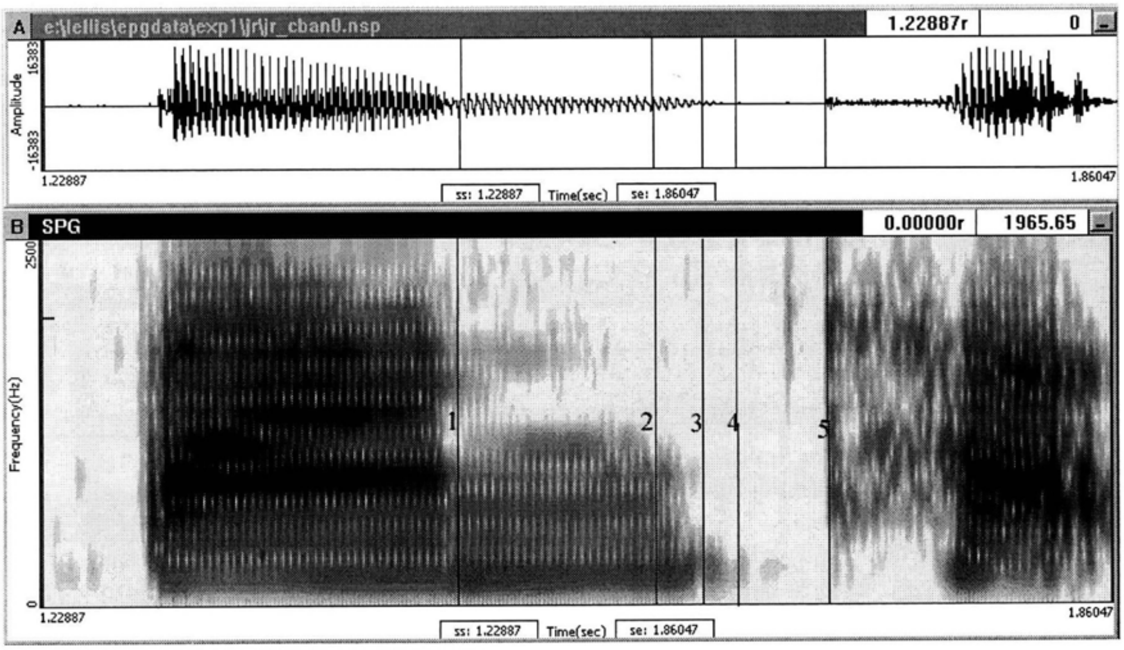
Subject J: careful speech

Subject J: fast speech

Figure 3.16 (x)

(caption as for Fig. 3.16 (i) above)

While the most gross variation in the timing bars is between assimilations and non-assimilations, a number of fine grained variations in the coordination of targets can be identified from the displays for non-assimilations. The most common coordinatory outcome involves the production of an alveolar and velar stop double articulation (voiced) which comes to an end with the release of the alveolar constriction while voicing continues during velar-only closure until the voiceless phase for /k/ begins. An examples from careful speech can be found on Figure 3.16 (iii) subject C repetition 10. Examples from fast speech can be found on Fig. 3.16 (v) subject E repetitions 7; Fig. 3.16 (iv) subject D repetition 5. The waveform, spectrogram and EPG patterns for one of these variants in careful speech is shown in Figure 3.17(i) below.



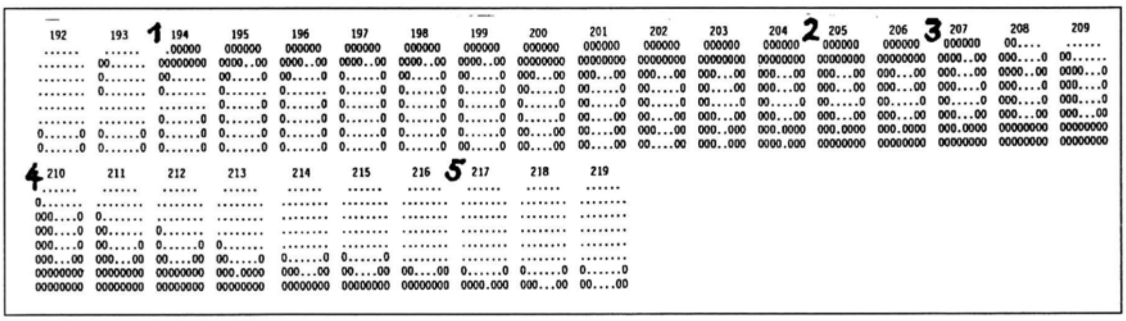


Figure 3.17 (i) –waveform, spectrogram and EPG patterns showing coordination of phonetic events for /n#k/ careful speech. Period of voiced double articulation, followed by alveolar release, followed by end of voicing. 1=onset of alveolar closure; 2=onset of velar closure; 3=release of alveolar; 4=end of voicing; 5=release of velar closure.

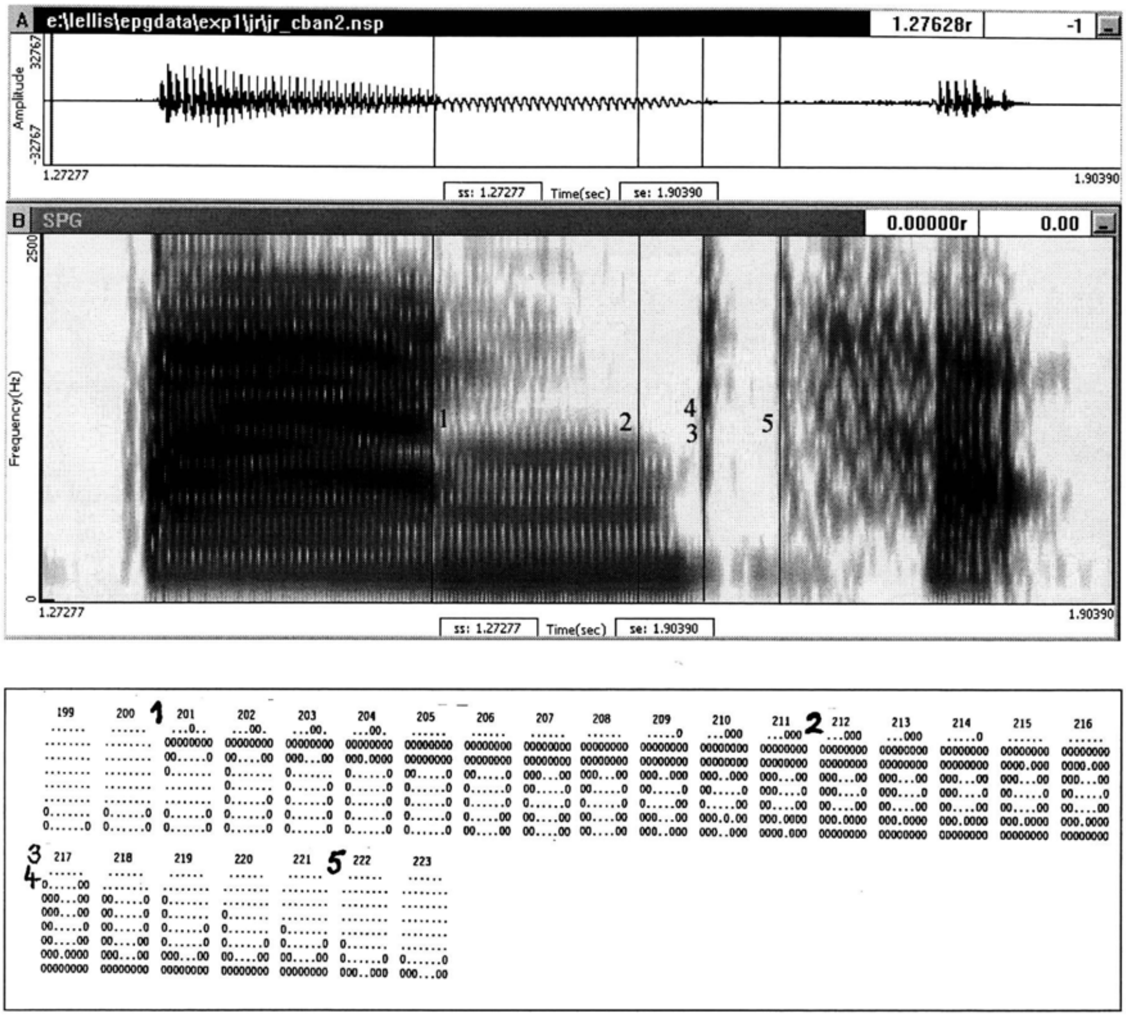


Figure 3.17 (ii) – waveform, spectrogram and EPG patterns showing coordination of phonetic events for /n#k/ careful speech. Period of voiced double articulation, followed by simultaneous alveolar release and end of voicing. 1=onset of alveolar closure; 2=onset of velar closure; 3=release of alveolar; 4=end of voicing; 5=release of velar closure.

An example of a variation on the most common coordinatory (overlap) strategy is when the release of the alveolar closure coincides with the end of voicing. Figure 3.17 (ii) shows a careful speech example of this.

For assimilations, however, the usual articulatory events are thus: onset of vowel; onset of stop closure /ŋ/; end of voicing; release of velar closure. The only variation on this is when there is no voiceless phase of velar closure. The velar release and the end of voicing are simultaneous. Subject B produced three of these in fast speech, as described in section 3.0 (see Figure 3.16 (ii), fast speech repetitions 2, 4 and 5). Figure 3.17(iii) shows the waveform, spectrogram and EPG patterns for one of these tokens.

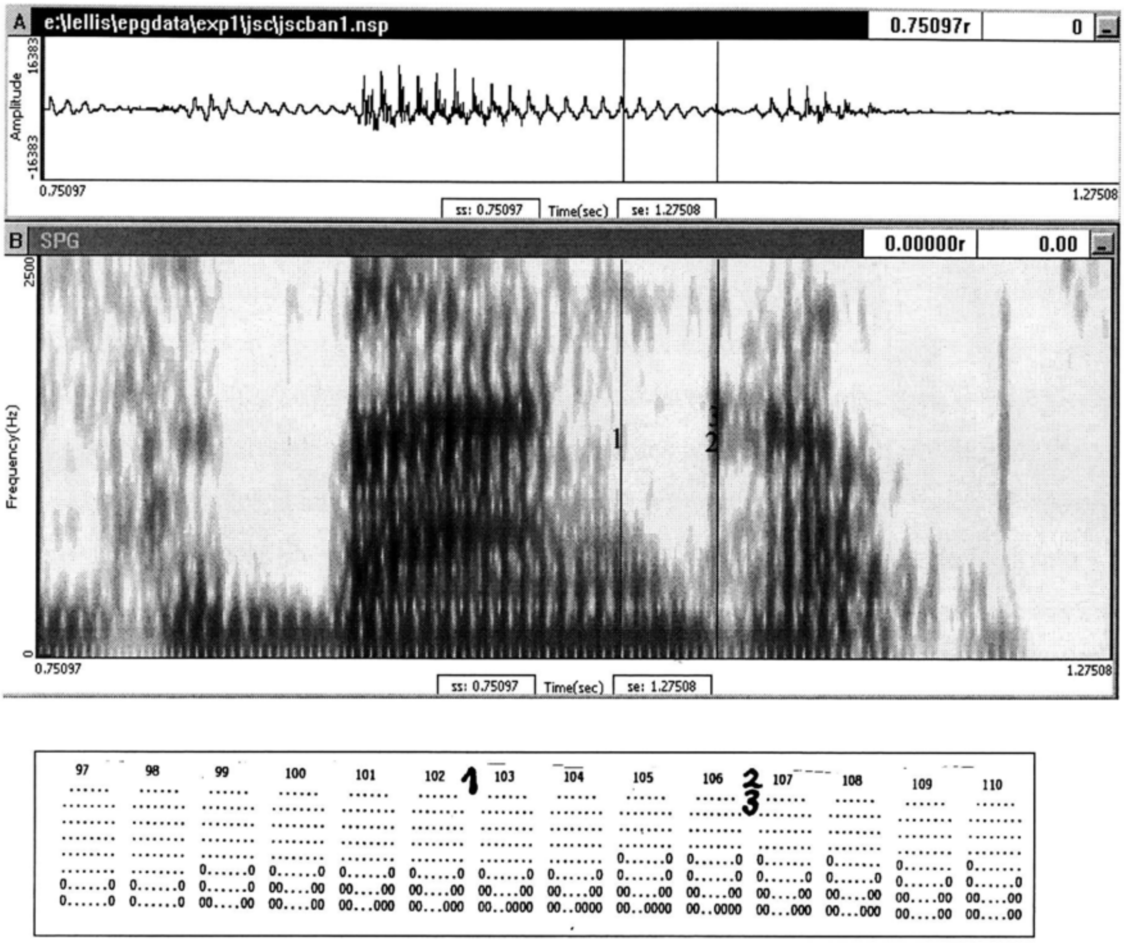


Figure 3.17(iii) – waveform, spectrogram and EPG patterns showing coordination of phonetic events for assimilated /n#k/ fast speech. There is no period of voiceless velar closure for /k/, release of velar closure and end of voicing are simultaneous. 1=onset of velar closure; 2=release of velar; 3= end of voicing

It should be noted here that the /n#k/ fast speech timing bars for subjects A and B are very variable relative to other subjects. We have seen that these subjects produce spatially different realisations of the alveolar to velar sequence, i.e. full alveolar, partial closure on the alveolar ridge, residual alveolar, complete assimilation, but they also vary more in timing aspects such as the coordination of the onset of voicelessness relative to oral stop closure and release events and overall duration of the sequence. The release of the velar in repetition 2 of subject A's fast speech timing bars is 100 ms earlier than that of the other complete assimilation, repetition 5. This difference is not found in any other subjects' fast speech data. The timing bars for subject B show considerable variability also. As already described, Repetition 2 amongst others shows no voiceless phase of oral closure, [baŋ gʌts], and repetitions 9 and 10 show differences in release of alveolar and velar closure timing.

3.2.1 Non-assimilation types

It became clear that within the category of 'non-assimilation' further distinctions between tokens could be made on the basis of the timing bar displays. Coordinatory outcomes can take three forms: 'overlap' (of alveolar and velar closure phases manifested as a period of double articulation EPG closure patterns), 'simultaneous' (simultaneous alveolar release and velar closure. N.B. not simultaneous alveolar and velar closure) and 'serial-ordering' (alveolar constriction released prior to velar closure). Further variants are noted (more details below) when location of onset of voicelessness is taken into account. An important overall finding was that release of the alveolar stop infrequently coincides with the offset of voicing for the voiceless phase of oral closure, a situation which does not fit with the discrete categories of segmental notation. Figure 3.18 shows EPG examples of each of the three non-assimilation types mentioned above. Fig. 3.18 (i) typifies the overlap type with a period of simultaneous alveolar and velar double articulation (frames 137-141 capture this), Figure 3.18 (ii) is an example of the simultaneous type where alveolar release and the completion of velar closure occur together at frames 227-228, Figure 3.18 (iii) is an example of the serial ordering type where velar closure at frame 174 follows alveolar release at frame 172.

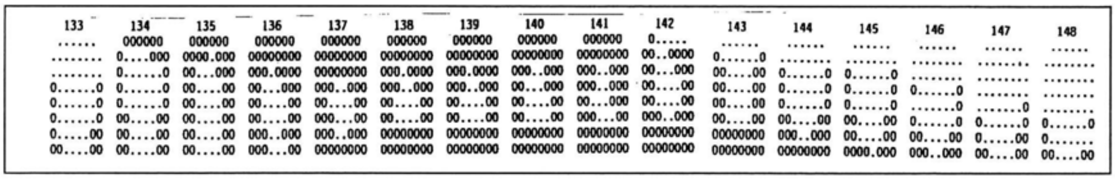


Figure 3.18 (i) non-assimilation 'overlap' type: Subject E fast speech

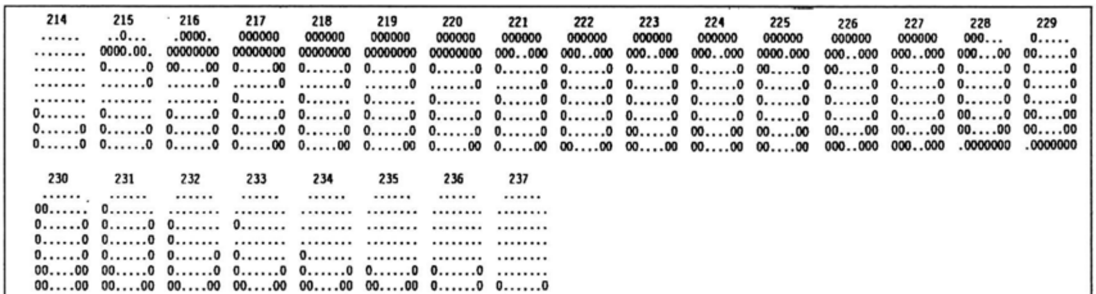


Figure 3.18 (ii) non-assimilation 'simultaneous' type: Subject H careful speech

```
  164      165      166      167      168      169      170      171      172      173      174      175      176      177      178      179
 ......   .0000.   .00000   .00000   .00000   .00000   .00000   .0000.   .000..   ......   ......   ......   ......   ......   ......   ......
........ 000.0000 000.0000 000.0000 000.0000 000.0000 00...000 00....00 00.....0 00...... 0....... ........ ........ ........ ........ ........
........ 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 ........ ........ ........ ........
........ .......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 .......0 .......0 .......0 ........ ........
........ .......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0
0......0 0......0 0......0 0......0 0......0 0......0 0......0 0......0 00.....0 00....00 00....00 00....00 00....00 0.....00 0.....00 0.....00
0......0 0......0 0......0 0......0 0......0 0.....00 0.....00 0.....00 0.....00 00....00 00....00 00....00 00....00 00....00 00....00 00....00
0......0 00.....0 00....00 00....00 00....00 00....00 000..000 000..000 000..000 000..000 00000000 00000000 00000000 00000000 00000000 00000000

  180      181      182      183      184      185      186
 ......   ......   ......   ......   ......   ......   ......
........ ........ ........ ........ ........ ........ ........
........ ........ ........ ........ ........ ........ ........
........ ........ ........ ........ ........ ........ ........
.......0 ........ ........ ........ ........ ........ ........
0......0 0......0 0......0 0......0 0......0 ........ ........
00....00 00....00 0.....00 0.....00 0......0 0......0 0......0
00000000 000..000 000..000 000..000 00....00 00....00 00....00
```

Figure 3.18 (iii) non-assimilation: serial-ordering type: Subject A careful speech

Inasmuch as these non-assimilation distinctions can constitute strategy oppositions, Table 3.10 shows their frequency in careful and fast speech and Table 3.11 shows their distribution across individual speakers. 'o/lap' corresponds to the overlap type, 'sim.' the simultaneous type and 's/o' the serial-ordering type. The values in brackets after each non-assimilation type in Table 3.10 indicates the percentage of all non-assimilations only, for each type.

There were very few serially-ordered tokens overall and those that were present were produced mainly by subject A in careful speech. Overall, the change from careful to fast speech appears to involve a proportionally higher incidence of simultaneous types (63% of all non-assimilations produced in fast speech versus 50% of all non-assimilations produced in careful speech) and fewer overlap types (35% versus 46%). However, as with the incidence of assimilations versus non-assimilations in fast speech (Figure 3.1) overall trends are qualified by speaker-specific effects. Looking at fast speech non-assimilations strategies for individual speakers, Table 3.11 below shows that all subjects except E produce more simultaneous types compared to overlap types. For subjects B, C and D there were more simultaneous types produced in careful speech also, which suggests a preference for this coordination strategy. Subject E preferred overlap types in careful and fast speech. But subjects A, and to a lesser extent F, show something of a switch in strategy from overlap in careful speech to mainly simultaneous in fast speech. It appears that speech rate has more of an effect on intergestural control of this sort for some speakers than others.

Table 3.10 distribution of /n#k/ types (pooled data) for fast and careful speech

careful speech				**fast speech**			
non-assimilation			**assimilation**	**non-assimilation**			**assimilation**
o/lap	**sim.**	**s/o**		**o/lap**	**sim.**	**s/o**	
44 (46%)	48 (50%)	4 (4%)	4	15 (35%)	27 (63%)	1 (2%)	57

Table 3.11 distribution /n#k/ types for individual speakers A-J, fast and careful speech

	careful speech				fast speech			
	non-assimilation			assim.	non-assimilation			assim.
subject	**o/lap**	**sim.**	**s/o**		**o/lap**	**sim.**	**s/o**	
A	4	3	3	-	-	7	-	3
B	1	9	-	-	-	3	1	6
C	1	9	-	-	1	3	-	6
D	2	8	-	-	3	5	-	2
E	9	1	-	-	7	3	-	-
F	10	-	-	-	4	6	-	-
G	9	-	-	1	-	-	-	10
H	1	5	1	3	-	-	-	10
I	7	3	-	-	-	-	-	10
J	-	10	-	-	-	-	-	10

Figure 3.19 below shows the rate of speech at which the various *careful speech* non-assimilation types were produced (rate of speech was measured as the interval between onset of /a/ and end of friction for /s/ in *...ban cuts...*). This confirms that the type of non-assimilation (i.e. 'overlap', 'simultaneous' or serial-ordering') is not predictable from the rate of speech and brings out speaker preferences more clearly than Table 3.11. It is acknowledged, however, that the deliberate contrasting of the three non-assimilation types may, in some cases, be based on small timing differences, that may, for instance, be indiscernible on EMA records. Figure 3.21 gives an indication of these timing differences. A separate variance t-test confirmed that the speech rates at which all 'overlap' and 'simultaneous' tokens were produced are not significantly different (p=0.9167). Both samples had a normal distribution. Figure 3.20 shows the speech rate for *fast speech* non-assimilation types (there were fewer non-assimilations in fast speech with subjects G-J producing none at all). Again there appears to be no relationship between speech rate and non-assimilation type. Again there was no significant difference between the speech rates at which all overlap and simultaneous types were produced.

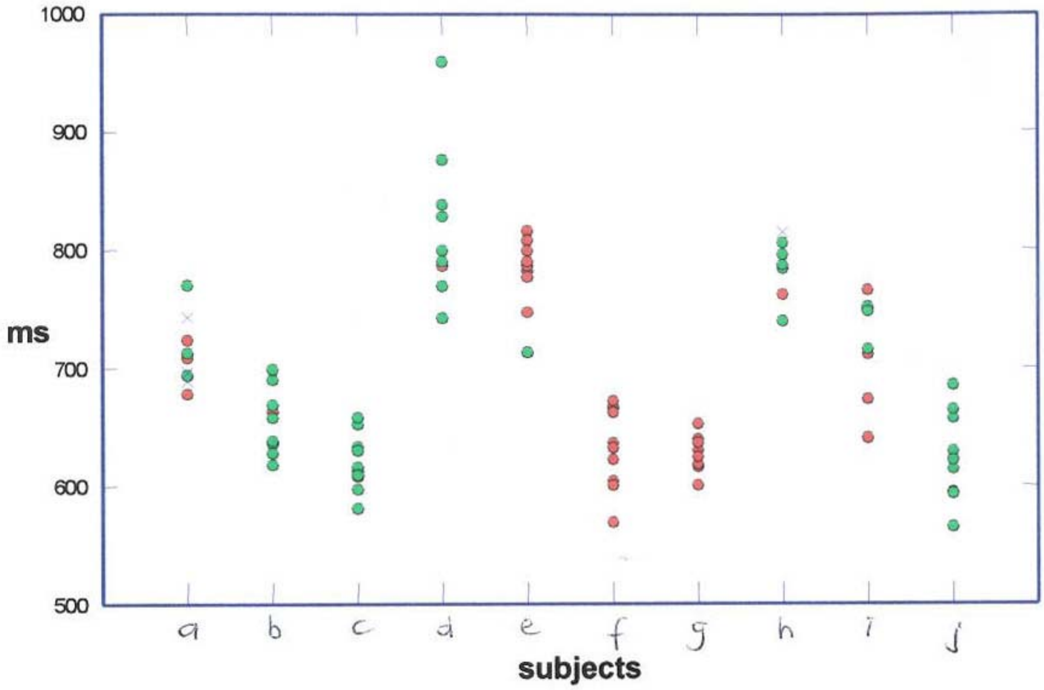


*Figure 3.19: scatterplot showing speech rate (measured as the interval between onset of /a/ and end of friction for /s/ in …*ban cuts*…) for* ***careful speech*** *non-assimilation types - overlap (red circles), simultaneous (green circles) and serial-ordering (blue crosses) for all subjects. 3 tokens are missing for subject H and 1 for subject G due to the fact they were assimilations*

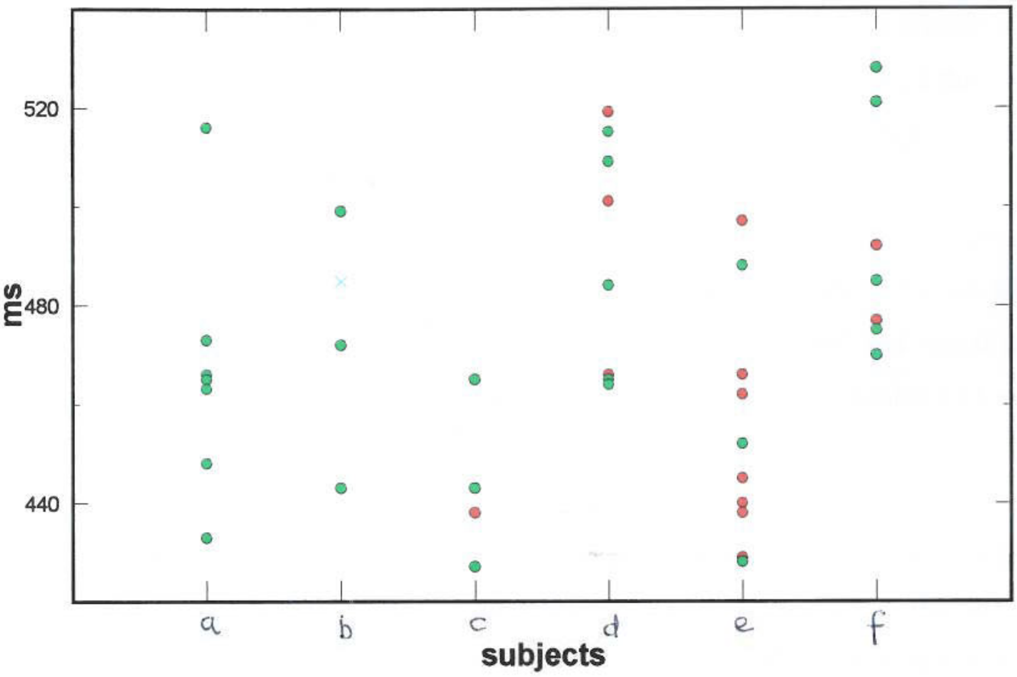


Figure 3.20 scatterplot showing speech rate (measured as the interval between onset of /a/ and end of friction for /s/ in ...ban cuts...) *for* ***fast speech*** *non-assimilation types - overlap (red circles), simultaneous (green circles) and serial-ordering (blue crosses) for subjects A-F*

The following sections will take each non-assimilation type and describe the variations that occur depending on the location of voicing and other factors. Close reference will be made to the timing bars, Figure 3.16 (i) – (x). After these three sections, a general comment will be made.

3.2.1.1 Non-assimilations: alveolar and velar stop closure overlap

The most common ordering of events for unassimilated /n#k/ is a double articulation period of simultaneous voiced alveolar and velar closure. Figures 3.21 (i) and (ii) are scatter plots showing the duration (ms) of double articulation for /n#k/ in careful speech and fast speech respectively. This was measured as the interval between the onset of velar closure and the release of the alveolar from the EPG data. Only non-assimilations are shown. In Figure 3.21 (ii) data for subjects G, H, I and J are missing since they always assimilated in fast speech and in Figure 3.19 (i) there are only four missing values (Subject H produced three assimilations and subject G produced one). Duration of double articulation 0-100ms relates to tokens that show stop closure overlap. But minus values on these scatterplots indicate the opposite of double articulation which is serial ordering. The greater the minus value the greater the interval between release of alveolar and onset of velar closure. And where the values are on or around zero, release and onset are simultaneous. Since the EPG sampling rate is 10 ms, a threshold of + or – 10 ms, was set (hence the horizontal lines on Figures 3.21 (i) and (ii)) so that tokens that appear on the graphs between -10 ms and +10ms, or on the lines, are assigned to the 'simultaneous' alveolar release and velar closure category. Tokens which involve overlap exceeding 10 ms in duration can be confidently identified as cases of overlap. The +/- 10 ms threshold defines the three coordinatory outcomes 'overlap', 'simultaneous' and 'serial-ordering' and the numbers of tokens in each category match up to the information in Tables 3.11 which was taken from the timing bars.

To find out if there is a correlation between the rate of speech at which tokens were produced and duration of double articulation a Spearman rank correlation was performed (the 'measure of double articulation' sample was not normally distributed). This showed that there was no correlation between these two measures for careful speech or for fast speech overall. The speakers who produced only non-assimilations in careful speech were tested individually to see if there was a correlation but none was found. Due to samples sizes of under 10, subjects who assimilated in either speech condition could not be tested for a correlation. The fast speech non-assimilations produced by Subjects E and F were tested for a correlation but again, no correlation was found. As speech rate increases the duration of overlapped alveolar and velar closure does not decrease and thus slower speech rates are not predictive of the presence of articulatory overlap.

Figures 3.22 (i) and (ii) show results for duration of double articulation as Figures 3.21 (i) and (ii) but the duration of double articulation is expressed as a percentage of a longer portion of the experimental sentence, from the onset of the vowel in *...ban cuts...* up to the end of /s/. Thus the data is normalised for variations in speech rate between and within speakers. What these scatterplots show is that the inter and intra-subject variability in the absolute values for both speech conditions are not the result of differences in speech rate but reflect true variation in the proportion of the utterance attributed to the double articulation. There seems to be no obvious constraint on duration of closure phase overlapping.

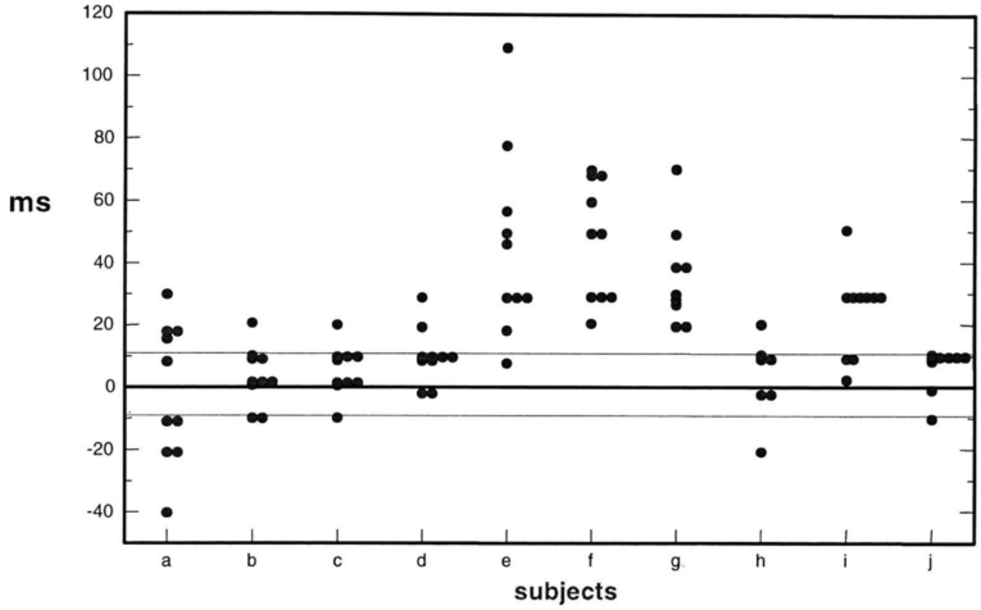


Figure 3.21 (i) /n#k/ careful speech - duration in ms of double articulation, all subjects, horizontal lines indicate the 10 ms EPG sampling rate 'margin of error'

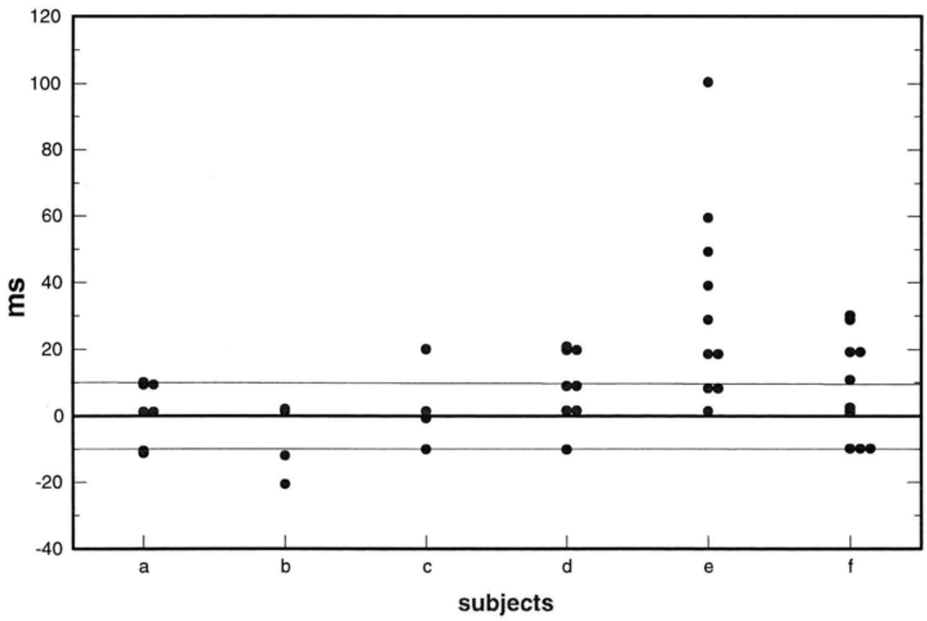


Figure 3.21 (ii) /n#k/ fast speech - duration in ms of double articulation, all subjects, horizontal lines indicate the 10 ms EPG sampling rate 'margin of error'

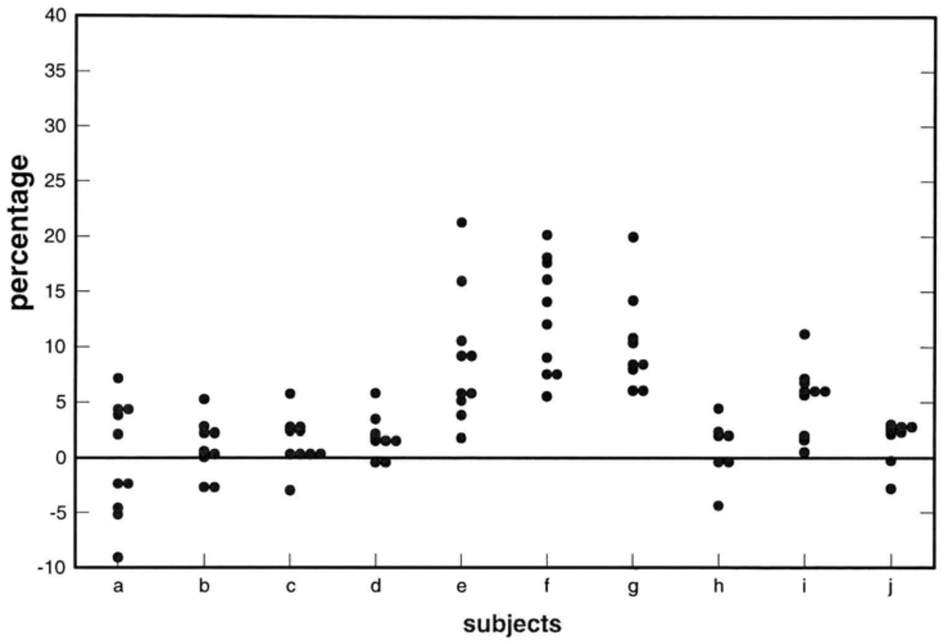


Figure 3.22 (i) /n#k/ careful speech – duration of double articulation expressed as a percentage of interval between onset of /a/ and end of friction for /s/...ban cuts...

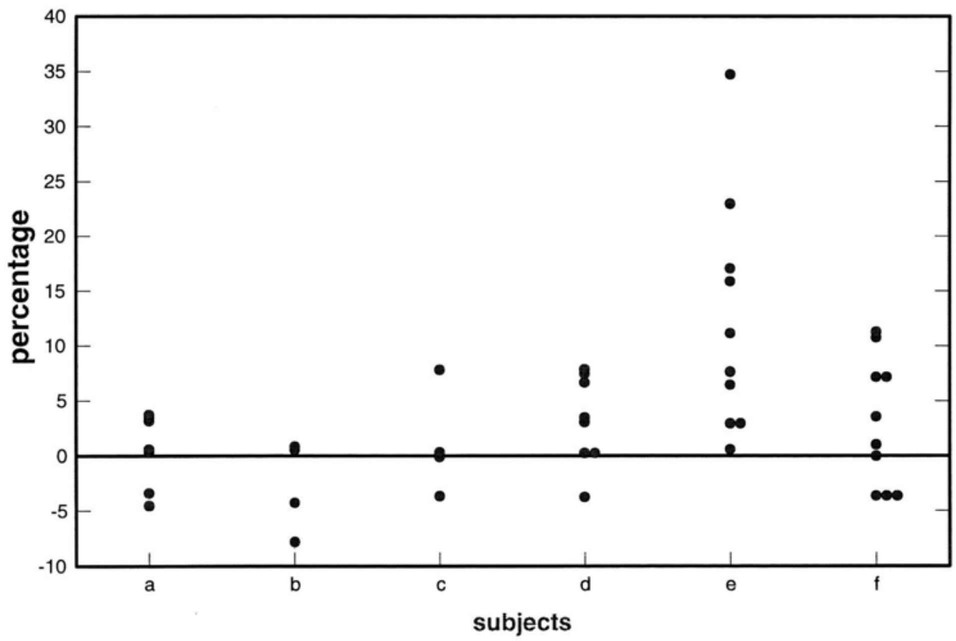


Figure 3.22 (ii) /n#k/ fast speech – duration of double articulation expressed as a percentage of interval between onset of /a/ and end of friction for /s/...ban cuts...

Observations based on the non-assimilation timing bars (Figure 3.16 (i) –(x)) show that there are several variations on the overlap strategy where there is still overlap but the relative location of closure and release movements and the onset and offset of voicing for the /n#k/ sequence is slightly different. These variations are listed below and for each, reference is made to specific timing bars as examples. The examples are not exhaustive. Where non-assimilation timing bars are not sufficiently clear to be either one type or another, the original EPG/acoustic data was checked. All overlap variations involve a period of double articulation during voicing. Variants (iii) and (iv) involve the continuation of a double articulation after voicing has ceased.

(i) double articulation comes to an end with the release of the alveolar constriction while voicing continues during velar-only closure until the voiceless phase for /k/ begins. Examples from careful speech: Figure 3.16 (i) subject A repetition 7; (viii) subject H repetition 4. Examples from fast speech: (i) subject A repetition 4; (v) subject E repetition 7.

(ii) release of alveolar closure and end of voicing are simultaneous with all other ordering the same (thus there is no voiced velar closure independent of alveolar closure). Examples from careful speech: Fig. 3.16 (ii) subject B repetition 3; (ix) subject I repetition 3. Example from fast speech: (iii) subject C repetition 8.

(iii) alveolar closure persists after the end of voicing and is released during the voiceless phase of velar closure for /k/. Examples from careful speech: Fig.3.16(v) subject E repetition 8 and 10; (iv) subject D repetition 9. Example from fast speech: (i) subject A repetition 5.

(iv) alveolar closure persists after the end of voicing and is released at the same time as the velar closure. Thus there is a period of voiced double articulation and a period of voiceless double articulation. There is only one example of this, it is produced by subject E and appears on Figure 3.16 (v) fast speech repetition 8.

Table 3.12 shows the distribution of these four overlap variants across all speakers for both speaking conditions. The percentages at the bottom indicate the proportion of the total *non-assimilations*, for either careful or fast speech, that are of each type i.e. (i), (ii), (iii) or (iv). These percentages can be checked against the percentage occurrence of all overlap types shown in Table 3.10 above.

Table 3.12 occurrence of non-assimilation overlap variants for individual speakers

	careful speech				fast speech			
subjects	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
A	4	-	-	-	-	-	-	-
B	1	-	-	-	-	-	-	-
C	-	1	-	-	-	1	-	-
D	1	-	1	-	2	1	-	-
E	6	1	2	-	3	3	-	1
F	7	3	-	-	3	1	-	-
G	9	-	-	-	-	-	-	-
H	1	-	-	-	-	-	-	-
I	6	1	-	-	-	-	-	-
J	-	-	-	-	-	-	-	-
*% of all non-assimilations	**36.5%**	**6.5%**	**3%**	**-**	**19%**	**14%**	**-**	**2%**

**percentage of either all careful speech non-assimilations or all fast speech non-assimilations (percentages for each condition do not add up to 100 until percentages for simultaneous and serial-ordering tokens are added in)*

3.2.1.2 Simultaneous alveolar release and velar closure

There are two variants on the 'simultaneous' strategy depending on the location of the end of voicing. Occurrence of 'simultaneous' tokens across all speakers for both speaking conditions can be see in Figures 3.21 (i) and (ii) These tokens fall between the - 10ms and +10ms EPG 'margin of error' on the graphs. The variants are listed below with some examples.

These variants are:

(i) release of alveolar and onset of velar is simultaneous but voicing is present during this. There is period of voiced velar closure before the onset of voicelessness. Careful speech examples: Figure 3.16 (viii) subject H repetition 9; Fig. 3.16(iii) subject C repetition 4. Fast speech example: Fig. 3.16 (ii) subject B repetition 7.

(ii) voicing ceases at the onset of velar closure. Three articulatory events occur simultaneously; two oral (alveolar release and velar closure) and one laryngeal (end of voicing). Careful speech examples: Fig. 3.16 (ii) subject B repetitions 7 and 10. Fast speech example: Fig. 3.16 (i) subject A repetition 3.

Table 3.13 shows the distribution of these two types across all speakers, careful and fast speech. The percentages at the bottom indicate the proportion of the total non-assimilations, for either careful or fast speech, that are of each type i.e. (i) or (ii). Again, these percentages can be checked against the percentage occurrence of all simultaneous types shown in Table 3.10 above.

Table 3.13 occurrence of non-assimilation simultaneous variants for individual speakers

	careful speech		**fast speech**	
subjects	(i)	(ii)	(i)	(ii)
A	3	-	5	2
B	4	5	3	-
C	8	1	2	1
D	7	1	3	2
E	1	-	3	-
F	-	-	6	-
G	-	-	-	-
H	4	1	-	-
I	3	-	-	-
J	10	-	-	-
*% of all non-assimilations	**42%**	**8%**	**51%**	**12%**

**percentage of either all careful speech non-assimilations or all fast speech non-assimilations (percentages for each condition do not add up to 100 until percentages for simultaneous and serial-ordering tokens are added in)*

3.2.1.3 Serially ordered alveolar release and velar closure

There are also two variants of 'serially ordered' tokens, again, depending on the location of the end of voicing. As for the other types of non-assimilation, occurrence of 'serially ordered' tokens across all speakers for both speaking conditions can be see in 3.21 (i) and (ii). On the timing bar displays (Figure 3.16) these can be identified by a gap between the end of the upper bar and the beginning of the lower bar during which time the tongue dorsum is completing velar closure after the tongue tip has come away from the alveolar ridge. These variants are:

(i) voicing is present as velar closure is formed. Careful speech examples: Figure 3.16 (i) subject A repetition 10; Fig. (viii) subject H repetition 10. Fast speech example: Fig. 3.16 (ii) subject B repetition 10.

(ii) voicing ceases at some point between alveolar release and velar closure with the result that the onset of velar closure is voiceless. There is one example of this on Fig. 3.16 (i) subject A careful speech repetition 8.

Table 3.14 shows the distribution of these two serial-ordering types across all speakers, careful and fast speech. The percentages at the bottom indicate the proportion of the total non-assimilations for either careful or fast speech that are of each type i.e. (i) or (ii).

Table 3.14 occurrence of non-assimilation serial-ordering variants for individual speakers

	careful speech		fast speech	
subjects	(i)	(ii)	(i)	(ii)
A	2	1	-	-
B	-	-	1	-
C	-	-	-	-
D	-	-	-	-
E	-	-	-	-
F	-	-	-	-
G	-	-	-	-
H	1	-	-	-
I	-	-	-	-
J	-	-	-	-
*% of all non-assimilations	**3%**	**1%**	**2%**	**-**

**percentage of either all careful speech non-assimilations or all fast speech non-assimilations (percentages for each condition do not add up to 100 until percentages for simultaneous and serial-ordering tokens are added in)*

From all three tables (3.12, 3.13 and 3.14), showing the occurrence of variants on overlap, simultaneous and serial-ordering types of non-assimilation, we can see that the first variant of each type '(i)' occurs more frequently than any of the other variants for each type, whether for careful or fast speech (77%). The common factor in all variant (i)s is that the alveolar closure is released before voicing ends and that there is a period of voiced velar closure (independent of alveolar closure) before voicing ends for /k/. In other words, release of the alveolar stop infrequently coincides with the offset of voicing, a situation which goes against the discrete categories of segmental notation. This is the case for all individual speakers (see Table 3.13, subject B, careful speech, for the only exception).

The following two sections deal with some further aspects of timing of the experimental sequence. Section 3.2.2 addresses the relationship between temporal latency of alveolar to velar closure (*cl1 – cl2*) and assimilation. Section 3.2.3 addresses the hypothesis that in the change from careful to fast speech, the duration of the vowel reduces in proportion to the duration of the non-assimilated consonant sequence. If this is the case then the domain at which rate-induced reduction takes place would be more likely the VCC than the CC unit.

3.2.2 Temporal latency and assimilation

Temporal latency refers to the interval between stop closures formed with the front and the back of the tongue for /n#k/. Figures 3.23 (i) and (ii) show the interval between onset of alveolar closure and the onset of velar closure (cl1-cl2) as a percentage of the duration between the onset of /a/ up to the end of /s/ in *...ban cuts...* for careful and fast speech respectively. The zero values in both graphs represent alveolar assimilations.

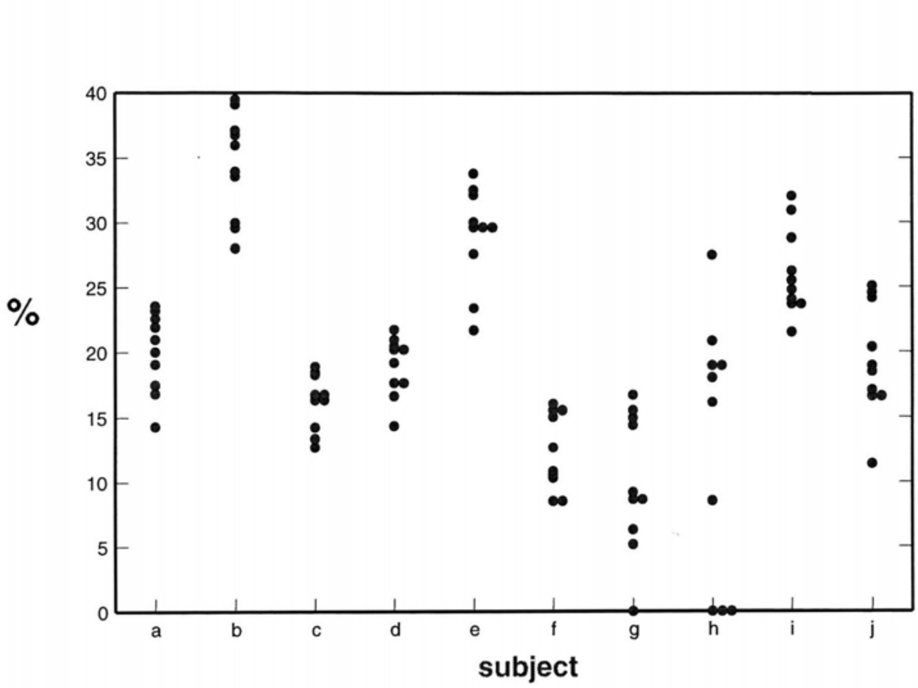


Figure 3.23 (i) /n#k/ careful speech – duration of cl1-cl2 expressed as a percentage of interval between onset of /a/ and end of friction for /s/ ...ban cuts...*Zero values=alveolar assimilations*

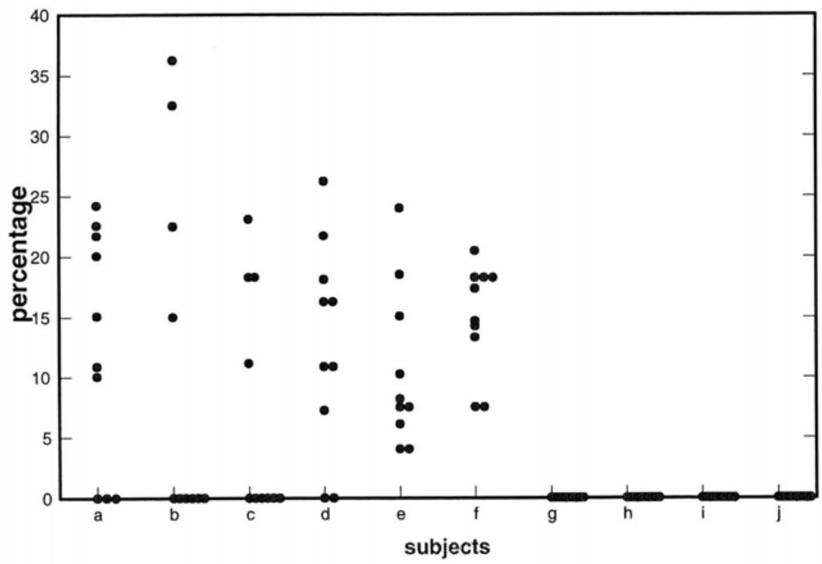


Figure 3.23 (ii) /n#k/ fast speech – duration of cl1-cl2 expressed as a percentage of interval between onset of /a/ and end of friction for /s/...ban cuts...*Zero values=alveolar assimilations (includes residual alveolars)*

In careful speech, the proportion of time taken up with cl1-cl2 varies considerably from speaker to speaker. An important factor that needs to be considered is whether a longer interval is due to the alveolar being released before the velar is formed. Figure 3.21 (i) indicates the occurrence of this type of sequencing and it appears that a relatively early alveolar release has no bearing on cl1-cl2 interval for careful speech. The intra-speaker percentage values for fast speech, however, are much more widely dispersed for fast speech compared with careful speech with less variation between speakers. A coefficient of variance value, enabling a comparison of variability spread, is given for all speakers for both speaking conditions in Table 3.15 below. High values indicate high variability.

Table 3.15 coefficient of variance values for % cl1-cl2: all subjects: careful and fast speech, non-assimilations

subject	**careful %**	**fast %**
A	15.3	31
B	11.8	34
C	13	34
D	12	37.2
E	13.3	60.7
F	24	29.8
G	38.1	-
H	29.3	-
I	13	-
J	22.4	-

Looking at Figures 3.23 (i) and (ii), there does not appear to be any obvious decreases in the percentage cl1-cl2 from careful to fast speech. The notion that as the cl1-cl2 interval becomes reduced to a certain degree, assimilation becomes likely, is not supported by the results. In careful speech, subjects F, G and H all have some percentage cl1-cl2 values of below 10% and yet only subject G and H produce assimilations. Similarly, in fast speech, Figure 3.23 (ii), there is no relationship between cl1-cl2 duration and assimilation. Subjects B and C contrast strongly with respect to percentage cl1-cl2 in careful speech. However, Figure 3.21 (i) shows that they are the same. Most of subject B and C's tokens fall within the 'simultaneous' ordering strategy.

To test the possibility that there is a correlation between percentage cl1-cl2 and speech rate (this time, rate as a continuous value and not just 'careful' or 'fast'), a Pearson Correlation was performed. For all careful speech data (subjects combined), samples for which were normally distributed, there was no correlation ($r^2 = 0.0697$). The scatterplot is shown in Figure 3.24 below. There were also no correlations on an individual-speaker basis. Similarly, there was no correlation between overall fast speech percentage cl1-cl2 and speech rate. Therefore, it appears that speech rate has no effect on this aspect of coordination for /n#k/.

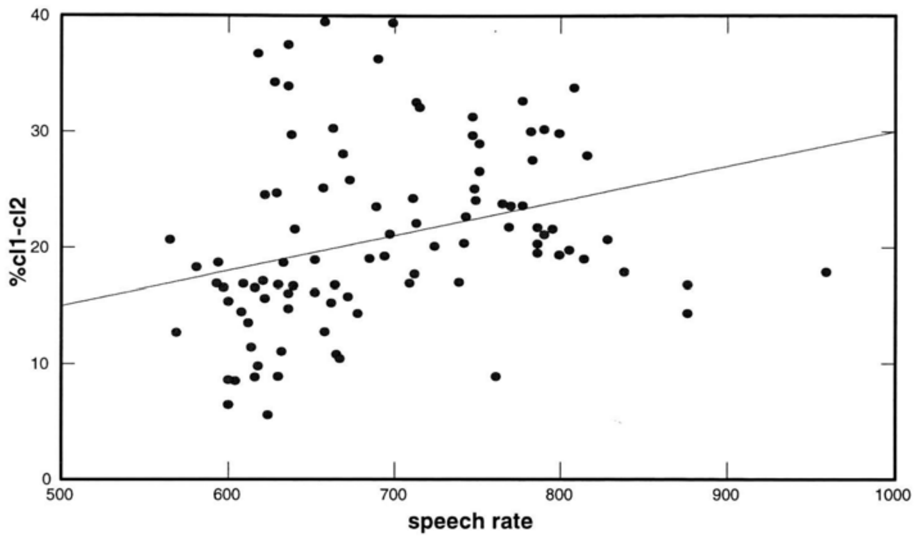


*Figure 3.24: absence of correlation between % cl1-cl2 and speech rate (interval between onset of /a/ and end of friction for /s/...*ban cuts...*) for all careful speech tokens (non-assimilations)*

3.2.3 Ratio of vowel to the rest of the sequence: non-assimilations

This section is intended to examine the possibility that in producing the $V_1C\#CV_2$ sequence ...*ban cuts*... at a faster rate, speakers maintain the same ratio of V_1 to the C#C cluster. All subjects reduce the absolute length of the vowel in fast speech, some more than others, as shown in Figures 3.25 (i) and (ii). Assimilations were taken out of both scatterplots.

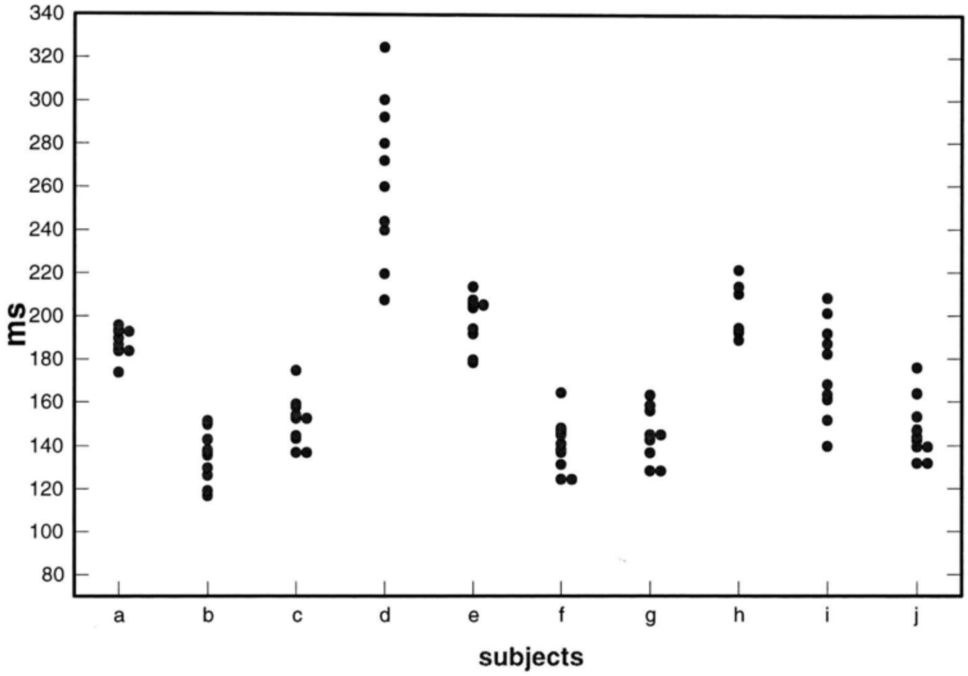


Figure 3.25 (i) absolute duration in ms of vowel in ...ban cuts..., *non-assimilations careful speech, all subjects*

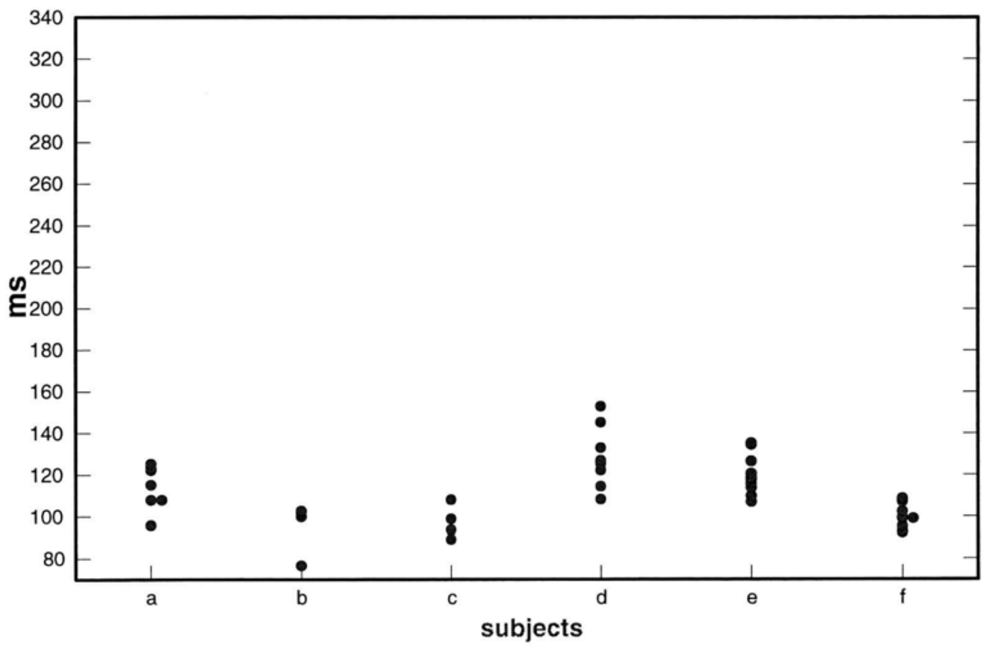


Figure 3.25 (ii) absolute duration in ms of vowel in ...ban cuts... *fast speech, subjects A-F*

As well as the duration of the vowel decreasing from careful to fast speech, the duration of the sequence, *cl1-re2*, was found to decrease. All subjects' careful speech *cl1-re2* durations fell between 120-300ms whereas in fast speech the slowest of these values was 140ms. There was an overlap of 20ms. The question to be asked now is whether, as the vowel gets shorter, the proportion of it to the C#C sequence is maintained. The 'C#C sequence' is measured here as the time between onset of the alveolar stop closure ('cl1', EPG-defined) and offset of velar closure ('re2', EPG-defined). Figure 3.26 (i) careful speech and Figure 3.26 (ii) fast speech show for all subjects the vowel as a percentage of the C#C sequence. A line is drawn across the graph at 100% so that any value above it indicates tokens where the vowel is longer than C#C. Assimilations are indicated as values at zero.

Combining all subjects, a separate variance t-test showed that there was no significant difference overall between the proportion of vowel to the non-assimilated consonantal sequence in careful speech compared to fast speech. It can be tentatively inferred from this that as the production of the experimental sequence speeds up, the vowel and the sequence in general both become shorter. However, it is clear from Figure 3.26 (i) and (ii) that there are significant differences for at least one subject, namely subject E. This was confirmed by a t-test which showed that there was a highly significant probability that the two samples were different, $p=<0.001$. For this speaker, then, at faster speech rates the vowel duration does not reduce in proportion to the duration of the experimental sequence, whereas for most speakers the relationship stays more or less the same. For one fast speech token produced by subject E the vowel is almost twice as long as C#C. This can be seen in the timing bars in Figure 3.16 (v), fast speech, repetition 4.

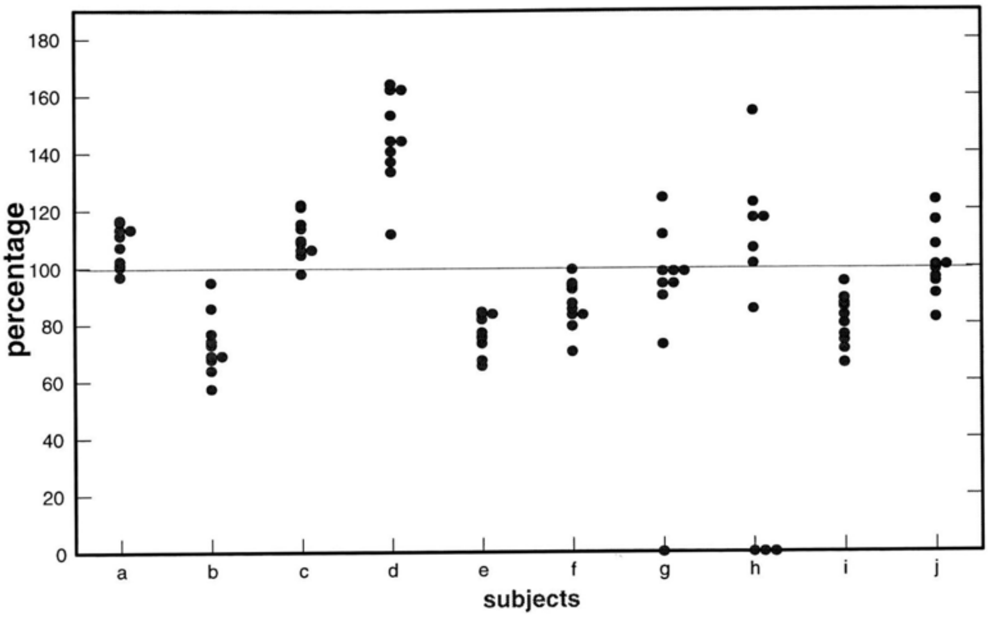


Figure 3.26 (i) vowel /a/ expressed as a percentage of cl1-re2 for all subjects careful speech...ban cuts...*(zero values = alveolar assimilations)*

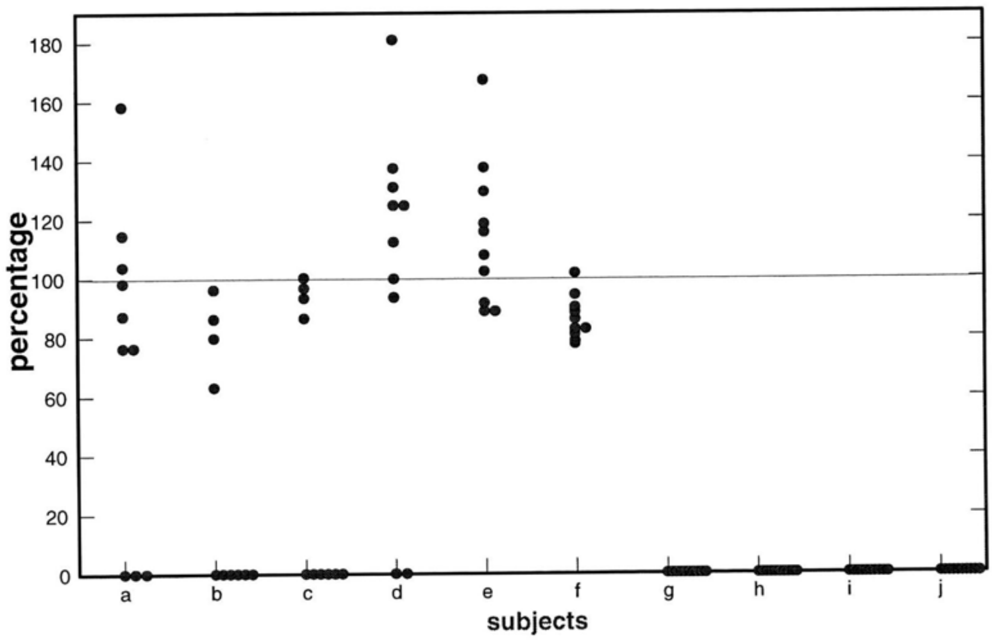


Figure 3.26 (ii) vowel /a/ expressed as a percentage of cl1-re2 for all subjects fast speech...ban cuts...*(zero values = alveolar assimilations)*

To summarise, a number of observations have been made from the results of the main EPG study. Firstly, assimilation is, overall, more frequent in fast speech compared to careful speech, although for some individual subjects, a faster rate does not motivate assimilation. There were very few assimilations in careful speech. In fast speech subjects either always assimilated, never assimilated or alternated between non-assimilation and assimilation. Of the subjects who alternated, two produced a continuum of /n#k/ reduction from non-assimilation to complete assimilation, while the other two produced either full alveolars or complete assimilations only. Comparisons made between the complete /n#k/ assimilations produced by those who always assimilated in fast speech and their lexical /ŋ#k/ showed that there was no place of articulation difference. However, for two of the four subjects in this group, there was a marginally significant difference between the duration of derived /ŋ/ and lexical /ŋ/ - lexical /ŋ/s were longer.

In terms of intergestural timing, it was found that the coordination of events for non-assimilations was variable with some evidence of individual-speaker preferences. Type of non-assimilation strategy (‘overlap’, ‘simultaneous’ or ‘serial ordering’) was not on the whole a function of speech rate. Scrutiny of the timing bars has revealed that more often than not, the alveolar stop is released before voicing ceases. No correlation was found between speech rate, as a continuous variable, and temporal latency between alveolar closure and velar closure for either careful speech ot fast speech. Occurrence of assimilation was not found to be related to duration of double articulation or to temporal latency.

CHAPTER FOUR

Follow-up EPG/EMA study: Introduction and Results

4.1 BACKGROUND TO THE COMBINED ELECTROPALATOGRAPHY (EPG) AND ELECTROMAGNETIC ARTICULOGRAPHY (EMA) STUDY

This section sets out the rationale for the re-recording of a subset of subjects from the EPG-only experiment, using the combined techniques of EPG and EMA. It also covers some methodological issues. Plans for the follow-up experiment were conceived on the basis of the results for the EPG-only experiment.

The follow-up combined method study is essentially a repeat of the original EPG-only study but with the simultaneous acquisition of additional kinematic information. The results of this study, however, are reported separately from the EPG-only experiment in order to draw attention to some specific research questions and to avoid any confusion which may arise from referring to sets of EPG data from different experiments. The subjects chosen to be re-recorded implemented the same /n#k/ assimilation strategy as they did for the EPG-only experiment.

Firstly, the rationale for re-recording some subjects will be described. Then, some issues and practical considerations arising from the use of the combined method will be briefly discussed. Following this some summarising research questions will be set out.

4.1.1 Rationale

A variety of speaker-specific strategies were identified from the original EPG experiment. Some subjects never assimilated, some always assimilated in a complete and categorical fashion and the rest optionally assimilated. A further sub-division of the group of optional assimilators could be made on the basis of their intra-speaker variability. Two subjects had a binary categorical opposition between non-assimilation and assimilation and two subjects produced a non-binary range of gradually reduced articulations. Assimilation then, *appears* to be categorical for some speakers while for others it is gradual. But, because EPG shows only the aspect of speech production which involves tongue contact with the hard palate and not movements away from and towards stricture, more subtle kinematic information (relating to observable properties of movements such as displacement and velocity) may prove the apparent categorical assimilation strategy to be gradual after all. Wright and Kerswill (1989) hypothesised that the tongue configuration for reduced coronal raising gestures was what caused the retracted EPG velar patterns they observed for some of their alveolar to velar assimilations: ‘as the tongue tip moves up towards the alveolar ridge, the blade and pre-dorsum become concave, which reduces the amount of lateral contact in the pre-velar area’. The stimuli which yielded these patterns was *...did gardens...*;*...lead got...*; and *...bed got...*Kühnert (1993) reported EMA evidence of residual alveolar gestures although she found these types occurred in the context of a more open and back vowel environment: /a, ɔ/ for German and /ɒ, ʌ/ for English, almost the reverse environment that Wright and Kerswill found retracted EPG patterns: ‘Thus only when the overall articulatory setting allows for it, in the form of a preceding low tongue configuration, can a residual raising movement of the tongue tip still be executed without affecting the palate.’ (p.268) However, Kühnert reported that all subjects produced these EMA-defined residual gestures at least some of the time.

Subjects who produced apparent complete assimilations in the EPG-only experiments were re-recorded to look again for gradience somewhere other than in the EPG trace. One speaker from the group of subjects who always assimilated in fast speech was selected, subject H, and one speaker from the group who varied between non-assimilation and complete assimilation was selected, subject D.

Because subjects A and B have already demonstrated phonetic gradience for alveolar to velar sequences in terms of tongue-palate contact it was not necessary to re-record these speakers to look for gradience. Only subjects C & D were re-recorded. But this does not mean to say that a more complete picture of underlying tongue movements for these speakers would be unenlightening. Evidence of residual tongue tip elevation for the complete assimilations they produced may lead us to speculate that phonetic identity between lexical velar-velar sequences and assimilated alveolar to velar sequences is never achieved. However, if this was not the case, then these forms can be regarded as extreme forms at one end of a reduction continuum thus avoiding the rather unparsimonious situation of evoking the application of a cognitive rule for these forms only. Another benefit of acquiring EMA data from the gradual assimilators would be the opportunity to observe overall tongue configurations for their EPG-defined residual alveolar articulations. If it is assumed that EPG-defined residuals are due to tongue tip raising for /n/, then the question arises of why the supporting tongue body gesture is not observed for the EMA-defined residuals. Kühnert (1993) did not show full EPG patterns in addition to her EMA data so that the occurrence of EPG evidence of residual alveolar could not be assessed. It was considered crucial for the present study to show full EPG patterns for sequences, allowing the opportunity to observe partial alveolar gestures on the hard-palate itself, of the type already identified in the EPG-only experiment.

4.1.2 Identification of residual alveolar gestures from EMA data

Following Kühnert (1993) residual alveolar gestures were identified by measuring maximum tongue tip height for assimilated /n#k/ for individual tokens. If any tokens are found to have vertical displacement in excess of that found for the /ŋ#k/ tokens those tokens will be considered partially assimilated. Since there is no specification of an alveolar target for /ŋ#k/, it is inferred that there will be little or no vertical tongue tip displacement. Any tongue tip raising present for /ŋ#k/ tokens will be considered merely biomechanical, as an automatic movement 'in sympathy' with the tongue back raising for the velar closure. It is solely the question of reduced coronal raising which will allow us to determine the phonological/phonetic status of this kind of assimilation for individual speakers.

4.1.3 The nature of EMA and EPG data

EPG is a well established technique for recording the timing and location of tongue contact with the hard palate during speech. While EPG provides information about tongue movement in the lateral plane and in the anterior region of the vocal tract, EMA gives us information on movement in the mid-sagittal plane and can give us a more complete picture of movement in the velar region. EPG cannot fully capture contact made at the junction between the hard and the soft palate where velar articulations are made. However, the advantage of EPG is that its highest resolution is for close approximations near the alveolar ridge, whereas EMA loses information here since tongue tip coils need to be located at least 1cm back from the tongue tip (so that there is minimal intereference with articulation). These two methodologies are an ideal combination for investigating the kinematic details of the production of consecutive front-back stop sequences. We have already seen that the analysis of EPG tongue-palate contact data independent of EMA data can go a considerable way in revealing the details of a variety of assimilatory processes, as previous studies have shown. Zsiga (1995) and Holst and Nolan (1995) successfully investigated the phenomenon of lexical and post-lexical palatalisation /s#j/ and alveolar to postalveolar assimilation /s#ʃ/ respectively. These studies, however, used EPG to look into same place of articulation assimilation which, unlike the present study, precluded the possibility of undetectable residual gestures. The analysis of EMA tongue trajectory data independent of EPG data, however, would not be sufficient to identify qualitatively different types of assimilation. It would not be possible to say with any certainty that one tongue tip raising trajectory resulted in target alveolar closure while another with comparable vertical displacement did not.

4.1.4 Summary EPG/EMA research questions

(i) Do subjects D and H produce reduced coronal gestures that leave no trace on the EPG patterns? If not, it shall be assumed that their assimilations can be attributed to the application of a cognitive rule.

(ii) Can the tongue configuration for this reduced coronal gesture be predicted from the EPG pattern (i.e. is this configuration likely to produce a retracted velar EPG pattern)?

4.2 RESULTS OF COMBINED EPG/EMA STUDY

Results of the follow-up EPG/EMA study will be presented in the following order. Firstly, fast speech EPG patterns for /n#k/ and control /ŋ#k/ produced by subjects D and H will be shown. Analysis of the EPG patterns served two purposes (i) to find out if each subject had replicated the assimilatory patterns they produced in the EPG-only experiment and thus applied the same assimilatory strategy and (ii) to identify those tokens which justified further investigation i.e. apparent completely assimilated /n#k/ tokens to look for coronal raising in the accompanying EMA data. Secondly, EMA data will be shown which compares the position of the articulators at the articulatory beginning of fast speech /n#k/ forms and of velar control sequences for the purpose of comparing tongue tip height. Thirdly, displays of tongue trajectories during production of individual fast speech /n#k/ and control /ŋ#k/ tokens will be shown. These will relate to the position displays in the previous section but will show the overall underlying tongue configuration (tongue tip, tongue body and tongue dorsum) for each token. Lastly careful speech assimilation data will be shown.

4.2.1 EPG patterns for fast speech /n#k/ and /ŋ#k/ subjects D and H

Results showed that each subject did in fact reproduce the assimilation strategy they applied in the EPG-only experiment. The full patterns are displayed in Figures 4.1 (i) for subject D and 4.1 (ii) for subject H. The EPG data is set out differently compared to previous EPG displays in this dissertation. Since the sampling rate for the EPG data in this follow-up experiment was set at 5ms, and thus there are twice the number of EPG frames for any excerpted sequence as there are for the data from the original EPG study which sampled every 10ms, only the frames leading up to closure and the first frame of closure itself are shown. This section of the utterance is sufficient to highlight the production of both an alveolar closure or a residual alveolar articulation (since residual alveolars are detected before velar closure in EPG data). Subject D produced 5 /n#k/ non-assimilations and 5 apparently complete assimilations. In the original EPG-only study this subject produced 8 /n#k/ non-assimilations and 2 apparently complete assimilations. Subject H produced 10 apparently complete assimilations in the EPG/EMA study as for the EPG-only study. This subject also produced one apparently complete assimilation in careful speech, this will be dealt with below.

It is necessary to note here the /n#k/ patterns for subject D. It has been reported that for each separate recording this subject produced maximally distinct forms of /n#k/, i.e./n/ or /ŋ/ in fast speech. While this result may not seem immediately obvious from a cursory inspection of the patterns yielded from the EPG/EMA recording there are good reasons why the patterns should be classified as either full non-assimilations or 'EPG-defined' complete assimilations and nothing in between. Subject D's /n#k/ repetitions 1, 3, 4, 5, 8 are non-assimilations. While these alveolar non-assimilations show less tongue-palate contact for the stop closure than most of those produced by the same subject in the EPG-only experiment (see contact patterns in Figure 3.4 (iv)), it is likely that this lack of full closure can be attributed to the presence of the EMA coil on the surface of the tongue tip/blade. Repetitions 1, 3, 4 and 5 all show mid-sagittal contact, the critical classificatory constituent of an alveolar described in section 2.1.5.1 of Chapter Two. Repetition 8, however, shows less alveolar contact than the other non-assimilations but in view of the spatial characteristics of the other non-assimilations, it will be considered a full alveolar stop.

The assimilations produced by subjects D and H were considered to be complete because none of these tokens showed lateral EPG contact further forward than the velar-velar controls. Figure 4.2 (i) and (ii) show all fast speech control sequences /ŋ#k/ produced by both subjects. Again due to the sampling rate, only the frames leading up to closure and the first frame of closure itself are shown.

For both subjects there were no 'retracted' velar stops for assimilated tokens. That is, there were no EPG velar patterns for assimilated /n#k/ which showed less velar contact than lexical velar tokens. Unfortunately the EPG data reduction methods available with EPG3 were not available in MATLAB the consequence of which is that some observations made and reported subsequently cannot be expressed numerically. For subject H there was a remarkable consistency in the fact that tongue-palate contact reached row 5 but not beyond this for 19 out of 20 tokens of /n#k/ and /ŋ#k/. For repetition 5 of /ŋ#k/ contact reached as far as row 3. There was no justification, then, for pursuing a correlation between tongue tip height and amount of velar closure for any of subject H's tokens. Unfortunately, we can do no more than speculate about this relationship for the assimilated /n#k/ velar retraction in the EPG-only study for subjects G, H and I.

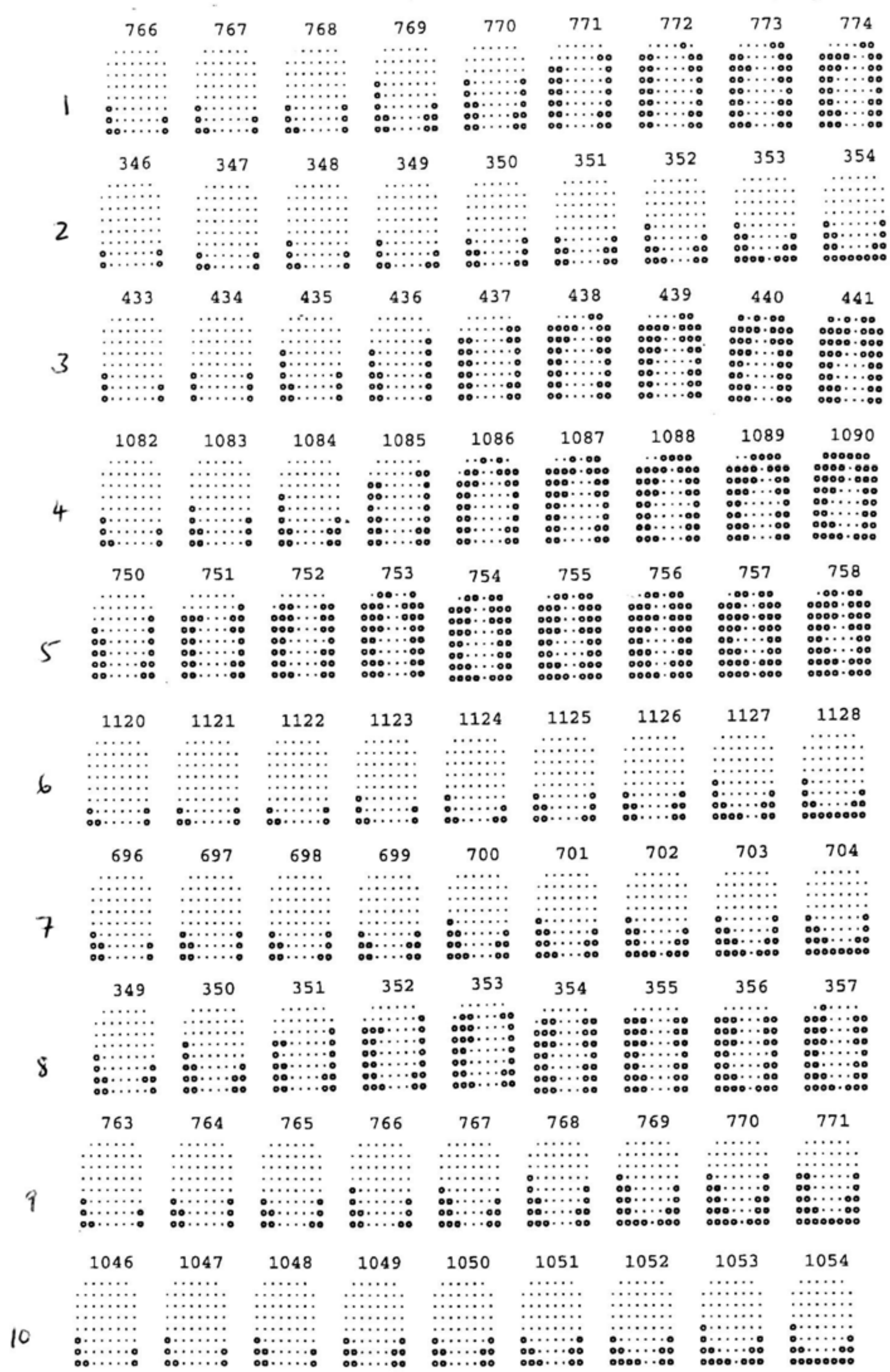


Figure 4.1 (i) fast speech /n#k/ EPG patterns from EPG/EMA experiment – subject D. All 10 (numbered) repetitions shown. Each line of patterns captures only the frames leading up to closure and the first frame of closure/maximum constriction.

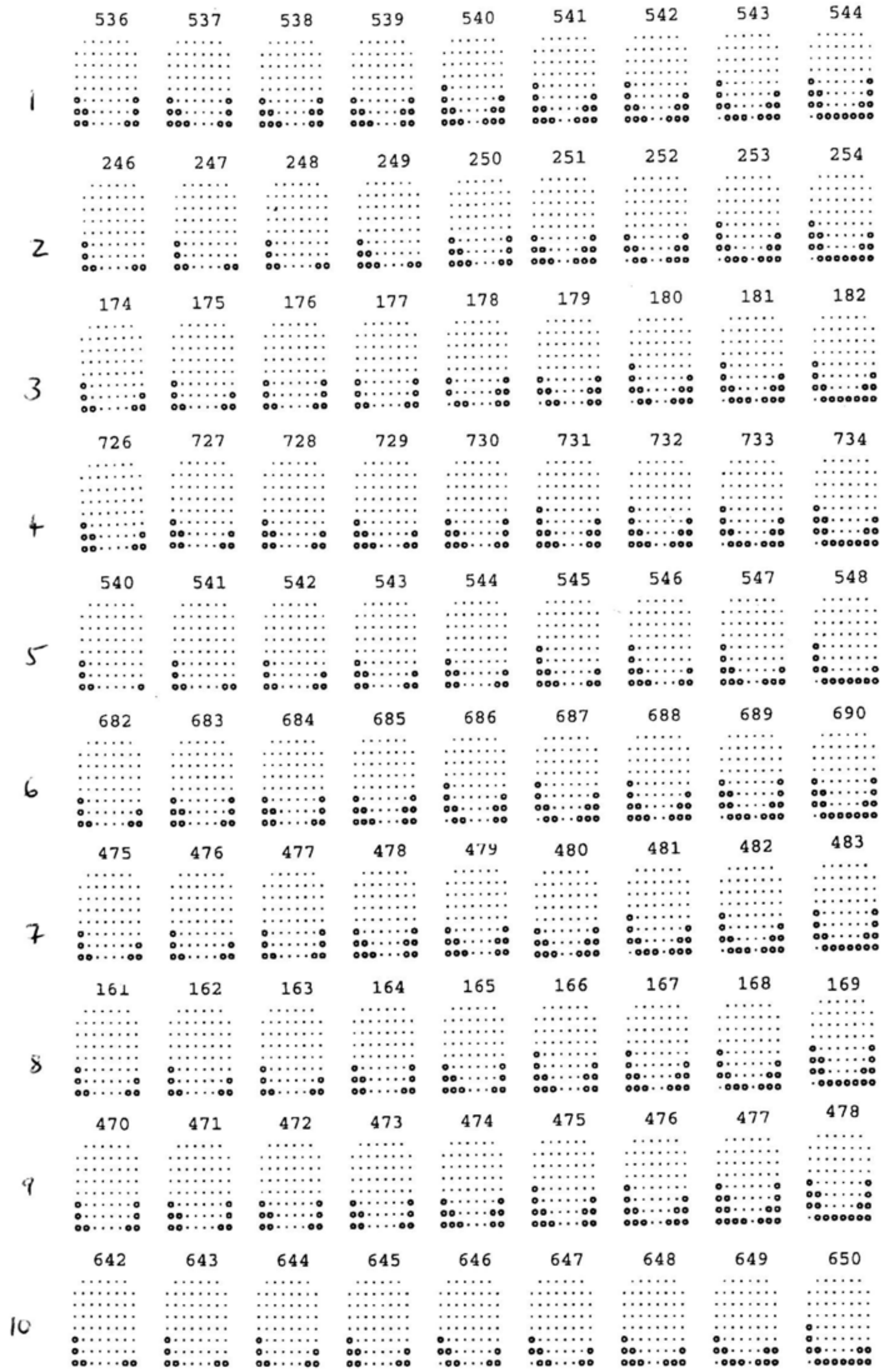


Figure 4.1 (ii) fast speech /n#k/ EPG patterns from EPG/EMA experiment – subject H. All 10 (numbered) repetitions shown. Each line of patterns captures only the frames leading up to closure and the first frame of closure/maximum constriction.

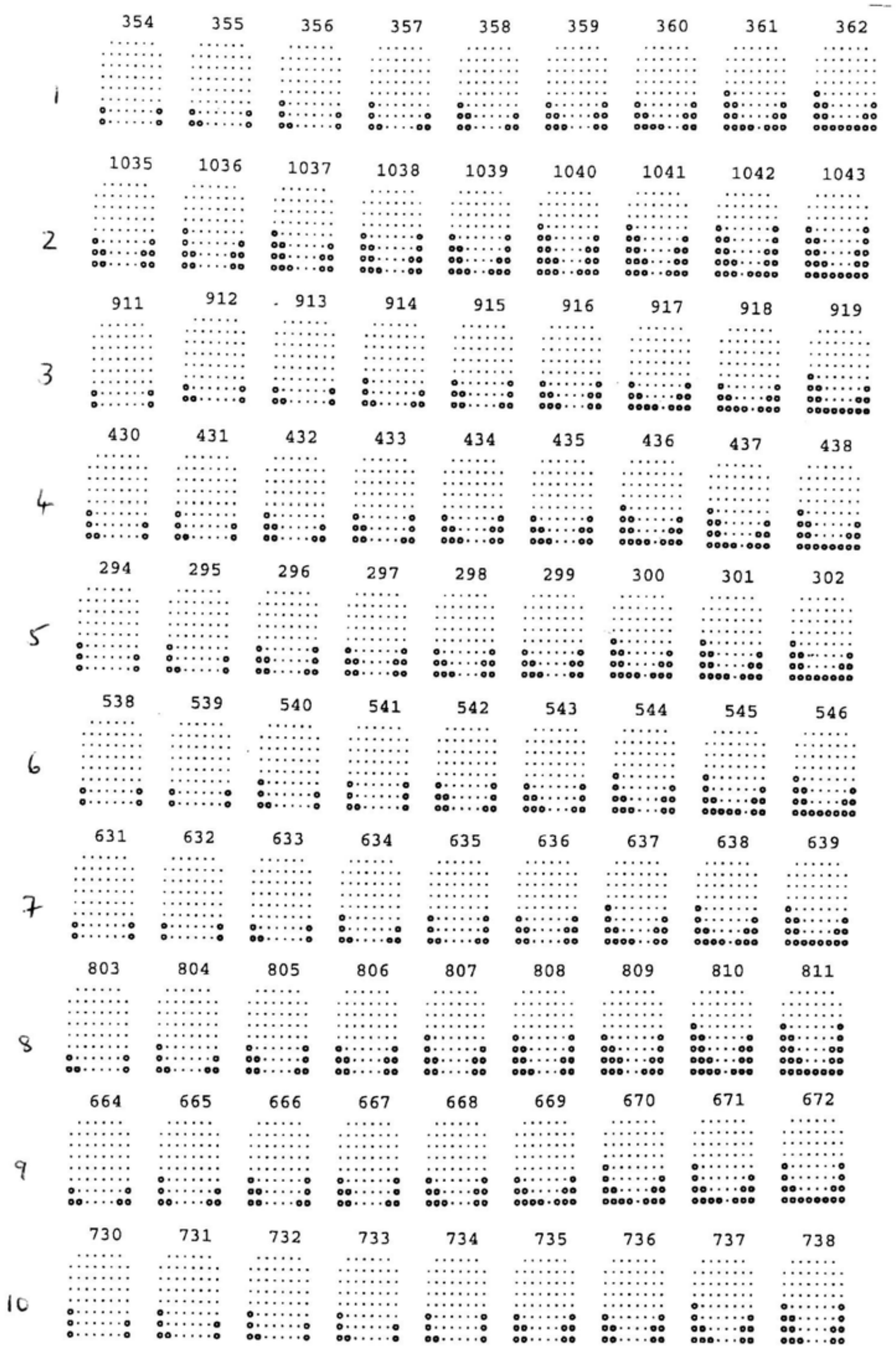


Figure 4.2 (i) fast speech control /ŋ#k/ EPG patterns from EPG/EMA experiment – subject D. All 10 (numbered) repetitions shown. Each line of patterns captures only the frames leading up to closure and the first frame of closure/maximum constriction.

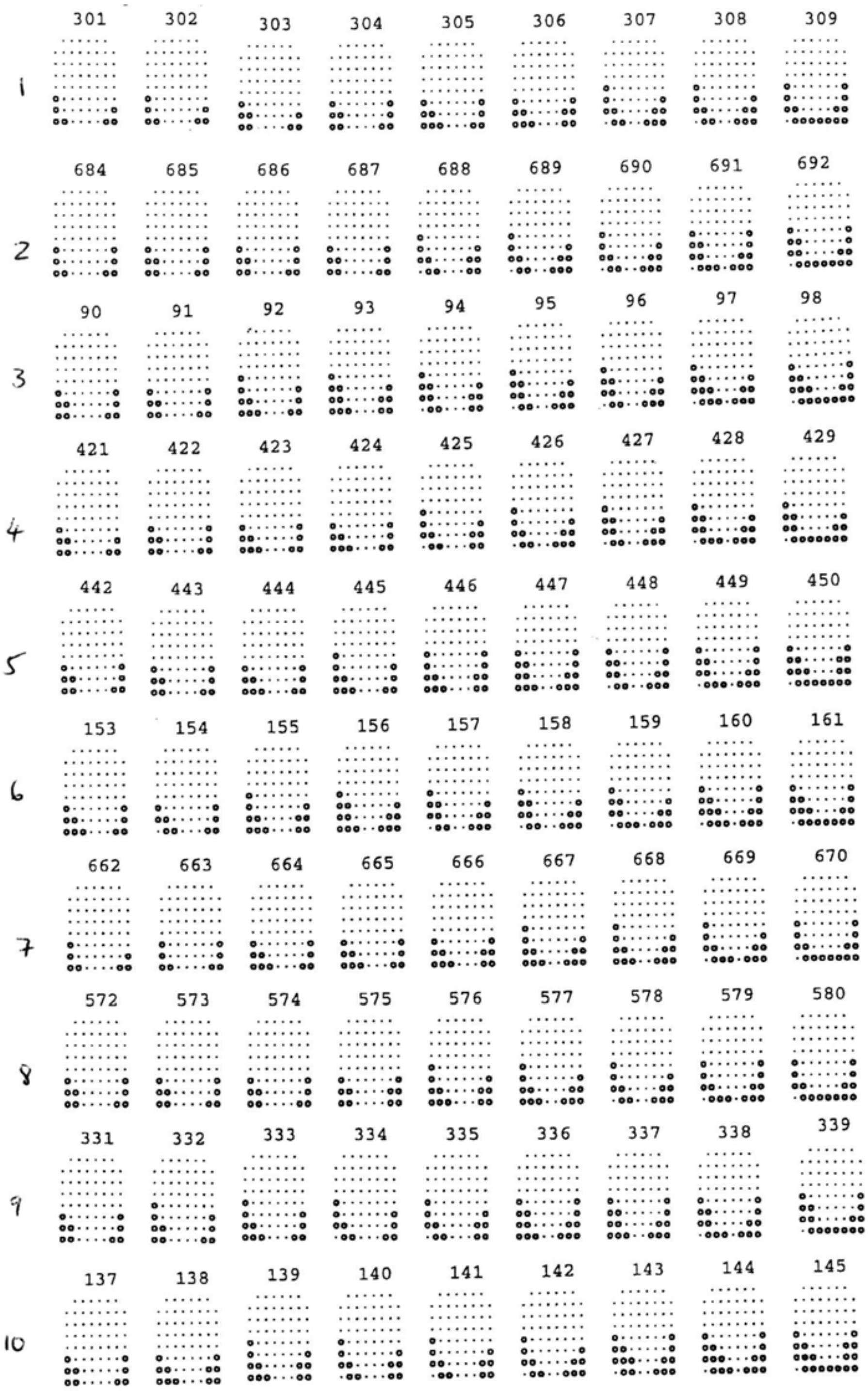


Figure 4.2 (ii) fast speech control /ŋ#k/ EPG patterns from EPG/EMA experiment – subject H. All 10 (numbered) repetitions shown. Each line of patterns captures only the frames leading up to closure and the first frame of closure/maximum constriction.

4.2.2 Articulatory positions for /n#k/ and /ŋ#k/ fast speech

In this section the position of the tongue tip and tongue dorsum at the articulatory beginning of all target /n#k/ tokens are compared with the positions at the beginning of all lexical /ŋ#k/ tokens.

Results for subject D are shown in 4.3 (i) and for subjects H are shown in 4.3 (ii). Position in x-axis and y-axis is shown in millimetres. The left hand cluster on each graph shows tongue tip positions and the right hand cluster shows tongue dorsum positions. Subject D's articulatory positions are plotted for non-assimilated alveolar sequences (numbering 5), apparent completely assimilated alveolar sequences (5) and all velar control sequences (10). For subject H, articulatory positions are plotted for apparent completely assimilated alveolar sequences (10) and velar control sequences (10). Subject H produced only 'complete' assimilations. It must be noted that tongue dorsum position is not shown on the graphs at the moment of maximum displacement but at the moment of maximum *tongue tip* displacement. Maximum tongue tip displacement defines the onset of the experimental cluster for non-assimilated /n#k/ tokens but for all assimilated /n#k/ tokens and control /ŋ#k/ sequences this is not so straightforwardly the case. If tongue tip displacement for an apparently complete /n#k/ token was sufficient to constitute a residual alveolar gesture then maximum displacement would indeed indicate the beginning of a cluster. But where there no residual raising gesture, the onset of the cluster would be at maximum tongue dorsum displacement.

In Figure 4.3 (i) maximum vertical displacement for the tongue tip is defined by the position for full alveolar closure for subject D's 5 non-assimilations. Here the tongue meets the unyielding boundary of the alveolar ridge. For this speaker, however, it is clear that the tongue tip cluster for the 'complete' /n#k/ assimilations overlaps with the tongue tip cluster for the underlying (lexical) velars. This means that vertical displacement for any assimilated /n#k/ sequence does not extend beyond that which normally accompanies a neutral velar control sequence for this speaker. No intermediate partial assimilation stage in between full alveolar stop closure and complete assimilation can thus be identified for subject D.

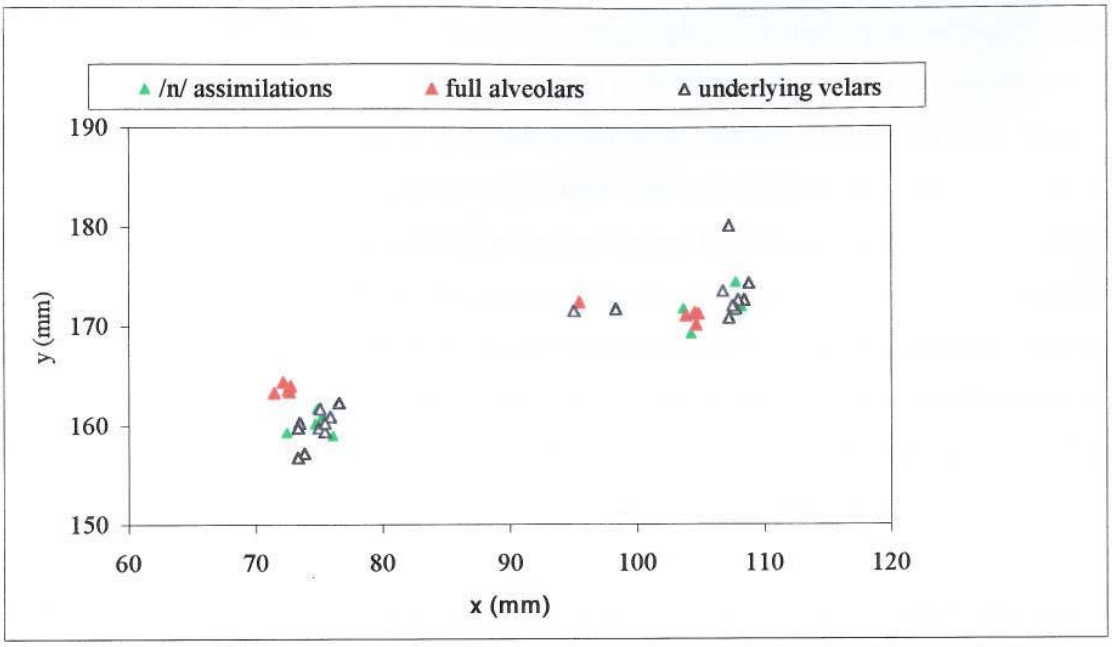


Figure 4.3 (i) Subject D: articulatory positions (mm) for tongue tip (left-hand cluster) and tongue dorsum (right-hand cluster) at the moment of maximum tongue tip displacement for all non-assimilated /n/ tokens (red triangles), assimilated /n/ tokens (green triangles) and all lexical /ŋ/ tokens(outlined).

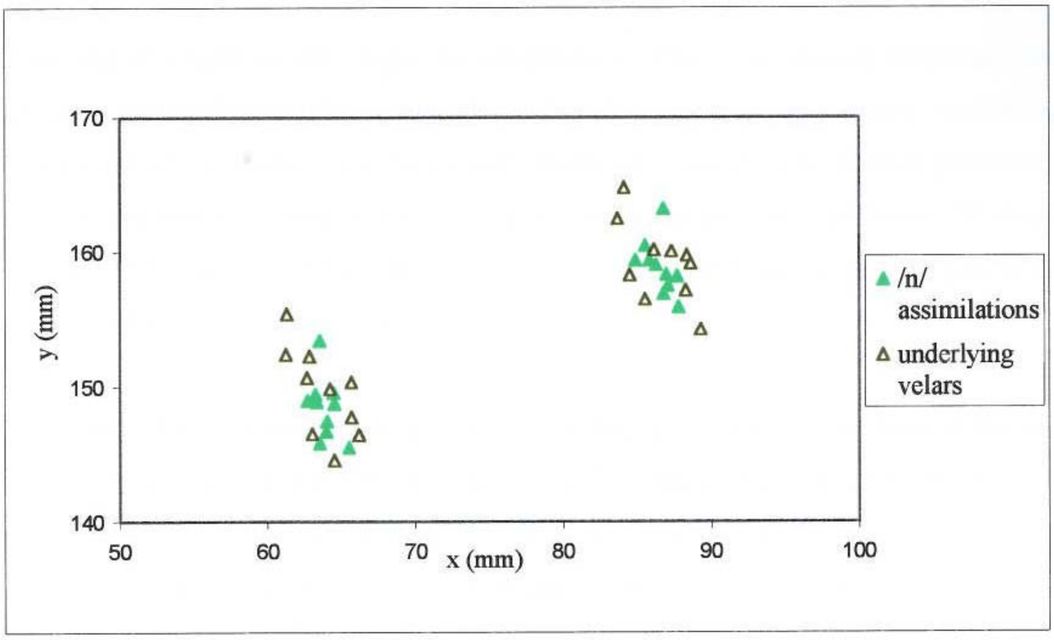


Figure 4.3 (ii) Subject H: articulatory positions (mm) for tongue tip and tongue dorsum at the moment of maximum tongue tip displacement for all assimilated /n/ tokens (green triangles) and all underlying /ŋ/ tokens (outlined triangles)

For Subject H, a similar picture emerges with overlap of tongue tip positions for assimilated /n#k/ and lexical /ŋ#k/ tokens. Since this subject produced only apparently complete assimilations for /n#k/, there is no indication on the graph where in the y-axis closure might occur. But it seems that the displacement of some underlying velar tokens is particularly extensive and must be very close to the position for actual contact with the alveolar ridge. Therefore, if the range of vertical tongue tip displacement for /ŋ#k/ is more extensive than it is for /n#k/ there can be no basis for interpreting *any* tongue tip raising movement associated with /n#k/ as evidence of a residual alveolar gesture. The range of tongue tip vertical displacement for all tokens produced by subject H is greater than for subject D even though subject H never achieves full alveolar closure. For subject H the distance between maximum and minimum tongue tip displacement for all tokens is11mm (144mm-155mm) and for subject D it is 7mm (157mm-164mm). This could be attributed to an overall more open jaw setting for subject H.

We have seen that for both speakers tongue tip position clusters for assimilated /n#k/ and control /ŋ#k/ overlap. But for both speakers (although to a lesser extent for subject D), the range of vertical displacement for control sequences is greater than that for assimilated /n/ forms, suggesting a possible constraint on movement variability for the latter. Tongue dorsum clusters for subjects D and H are also more constrained for the assimilated form of /n#k/ than for /ŋ#k/. It is worth noting here that the non-assimilation tokens produced by subject D show a tight clustering of tongue tip and tongue dorsum position. This is not entirely surprising since these articulations involve actual contact with the alveolar ridge and a subsequent ‘anchoring’ effect on the back of the tongue for velar closure. However, x-axis tongue dorsum position for one non-assimilation token (red triangle) could be considered an outlier at 95mm. With relation to the /n#k/ EPG patterns for this speaker, Figure 4.1 (i), there is no single EPG pattern which is markedly different from the others.

For subject H the dorsum and tongue tip clusters for /ŋ#k/ seem to vary more in the vertical dimension compared to the /n#k/ clusters, while for subject D, position variability seems to be comparatively extensive in the horizontal dimension. For the latter speaker, x-axis dorsum position for /ŋ#k/ varies markedly from 95mm-109mm, a considerable distance of 14mm. With reference to the EPG patterns, we can see that /ŋ#k/, Figure 4.2 (i) repetition 2 and 8 show a more fronted tongue back position in relation to all the other tokens. By contrast dorsum position for assimilated /n#k/ is more constrained.

4.2.3 Displays of tongue coil trajectories

The position graphs above show the x-y position of articulators for all tokens but it is not possible to relate tongue tip position for a single token to its corresponding tongue dorsum position and so gain an impression of the underlying tongue configuration for individual tokens. Subjects D and H's /n#k/ displays are presented in Figure 4.4 (i) and (ii) respectively and their /ŋ#k/ displays in Figure 4.5 (i) and (ii) respectively. These displays show the dynamic displacement trajectories for the tongue tip, tongue body and tongue dorsum for each individual token from the middle of the vowel /a/ up to the middle of the post-target sequence vowel /ʌ/ in fast speech*...ban cuts...* or *...bang comes...* A solid line linking all 3 coil trajectories for each token represents an interpolated tongue configuration either at maximum tongue tip or tongue dorsum displacement. For subject D, displays for non-assimilated /n#k/ tokens (numbering 5) show tongue configuration at maximum articulatory displacement of the tongue tip coil which corresponds to formation of full stop closure for /n/, i.e. the beginning of the consonant sequence. All other displays for this subject and all displays for subject H, including those of control sequences, will show tongue configuration at the moment of maximum articulatory displacement for the tongue dorsum coil. This is so because it has already been established from the articulatory position results in Figure 4.3 that the apparent completely assimilated /n#k/ tokens produced by both subjects do not in fact reveal residual alveolar gestures. It is assumed in these cases that the observed trajectory of alveolar movement is due to biomechanical movement 'in sympathy' with the prevailing tongue dorsum movement for velar closure. The control sequences, of course, have no alveolar specification and so the onset of the sequence is velar anyway. Therefore the onset of velar closure is taken to be the beginning of the sequence. It must be noted here that the tongue tip coil in the displays for subject D's /n#k/ tokens where a full alveolar closure is made (displays 1-5), never actually *appears* to reach the alveolar ridge shown as part of the displayed palate trace. This is probably because the position of that particular coil on the surface of subject D's tongue was just outside the area of the tongue tip/blade which actually made contact with the ridge. Subject H's trajectories do not match with the palate trace display because of an error in adding the palate traces to the data. It must also be pointed out that the trajectory display 5 for subject D /ŋ#k/ Figure 4.5 (i) is very unusual. The movement pathway for the tongue dorsum is very high and would seem to indicate that the dorsum coil had been loosened from the tongue surface. Neither the previous nor the following repetition produced by subject D appear to be similarly affected. This accounts for the outlier in the articulatory position graph in Figure 4.3 (i).

The individual trajectory displays of assimilated /n#k/ for both subjects show none of the 'tongue hollowing' which, as suggested by Wright and Kerswill (1989), may be characteristic of residual alveolars. For neither subject is the position of the tongue tip coil higher than the tongue body coil. It is, of course, possible that there is some hollowing between the tongue body and tongue tip coils unobservable from the displays. Furthermore, there appear to be no consistent difference in tongue configuration when both subject's assimilated /n#k/ displays are compared with their lexical /ŋ#k/ displays. There is, however, a difference in the fact that the solid line on subject H's displays tends to be straight whereas for subject D, the tongue body is typically higher.

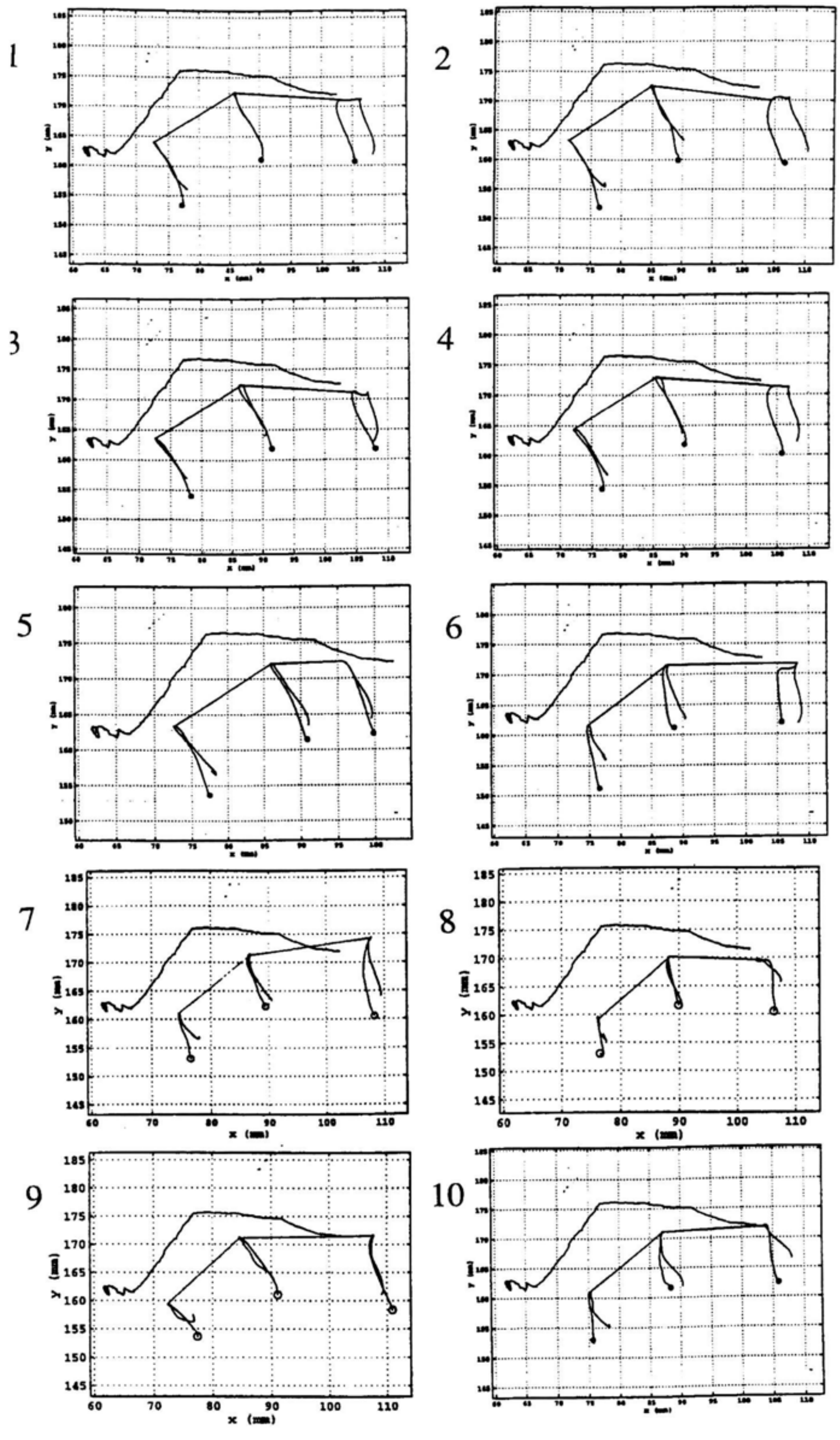


Figure 4.4 (i) Coil trajectory displays for individual fast speech /n#k/ tokens produced by ***subject D****. The non-assimilations are grouped as tokens 1-5, this does not reflect the order in which they were produced. Movement paths of the tongue tip, tongue body and tongue dorsum coils are shown from the mid-point of pre-experimental sequence /a/ up to the mid-point of the following /ʌ/. The solid line is an interpolation of tongue configuration either at maximum tongue tip displacement (for non-assimilations, here numbers 1-5) or at maximum tongue dorsum displacement (for assimilations numbers 6-10).*

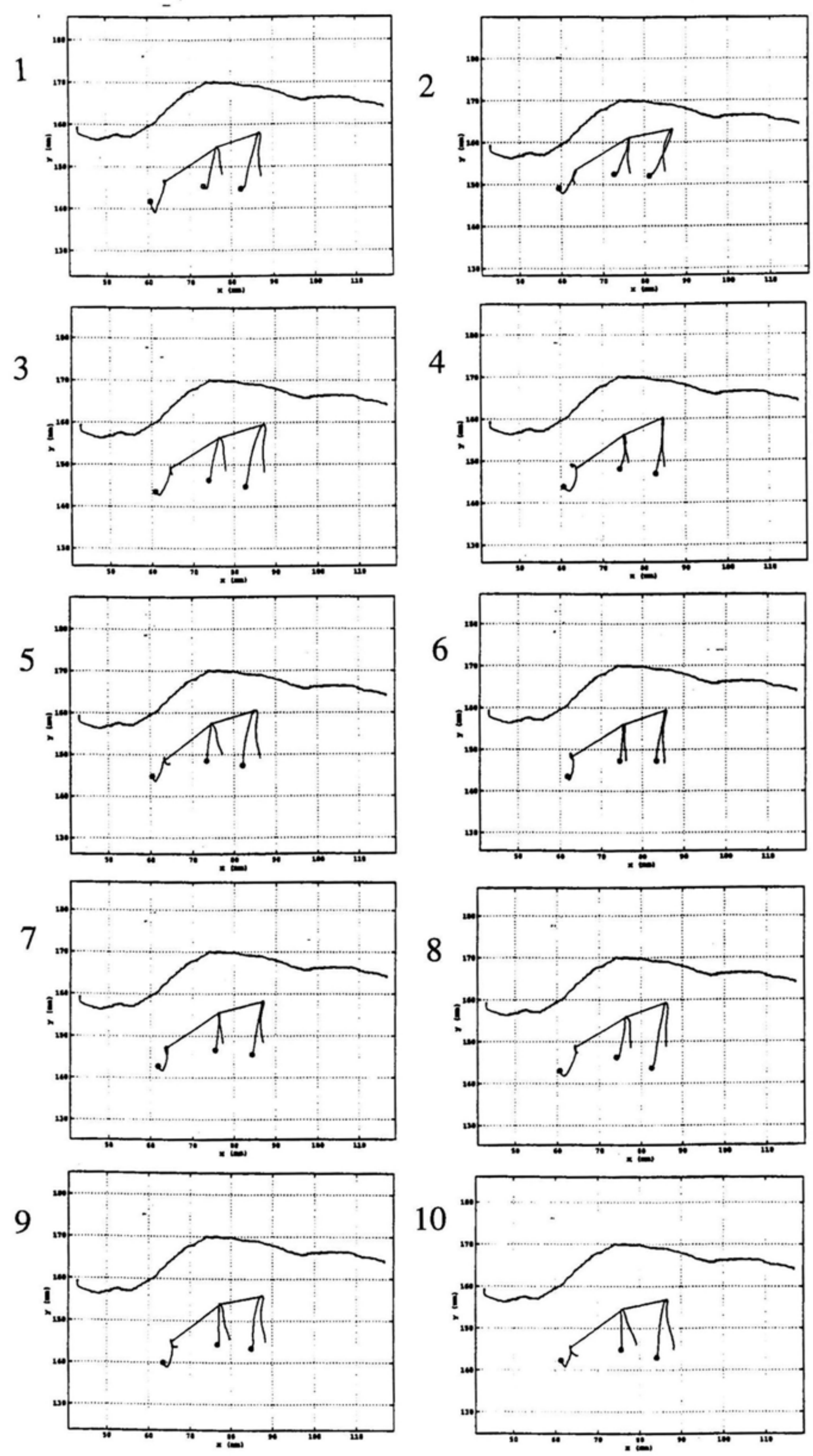


Figure 4.4 (ii) Coil trajectory displays for individual fast speech /n#k/ *tokens produced by* ***subject H****. Movement paths of the tongue tip, tongue body and tongue dorsum coils are shown from the mid-point of pre-experimental sequence /a/ up to the mid-point of the following /ʌ/. The solid line is an interpolation of tongue configuration at maximum tongue dorsum displacement.*

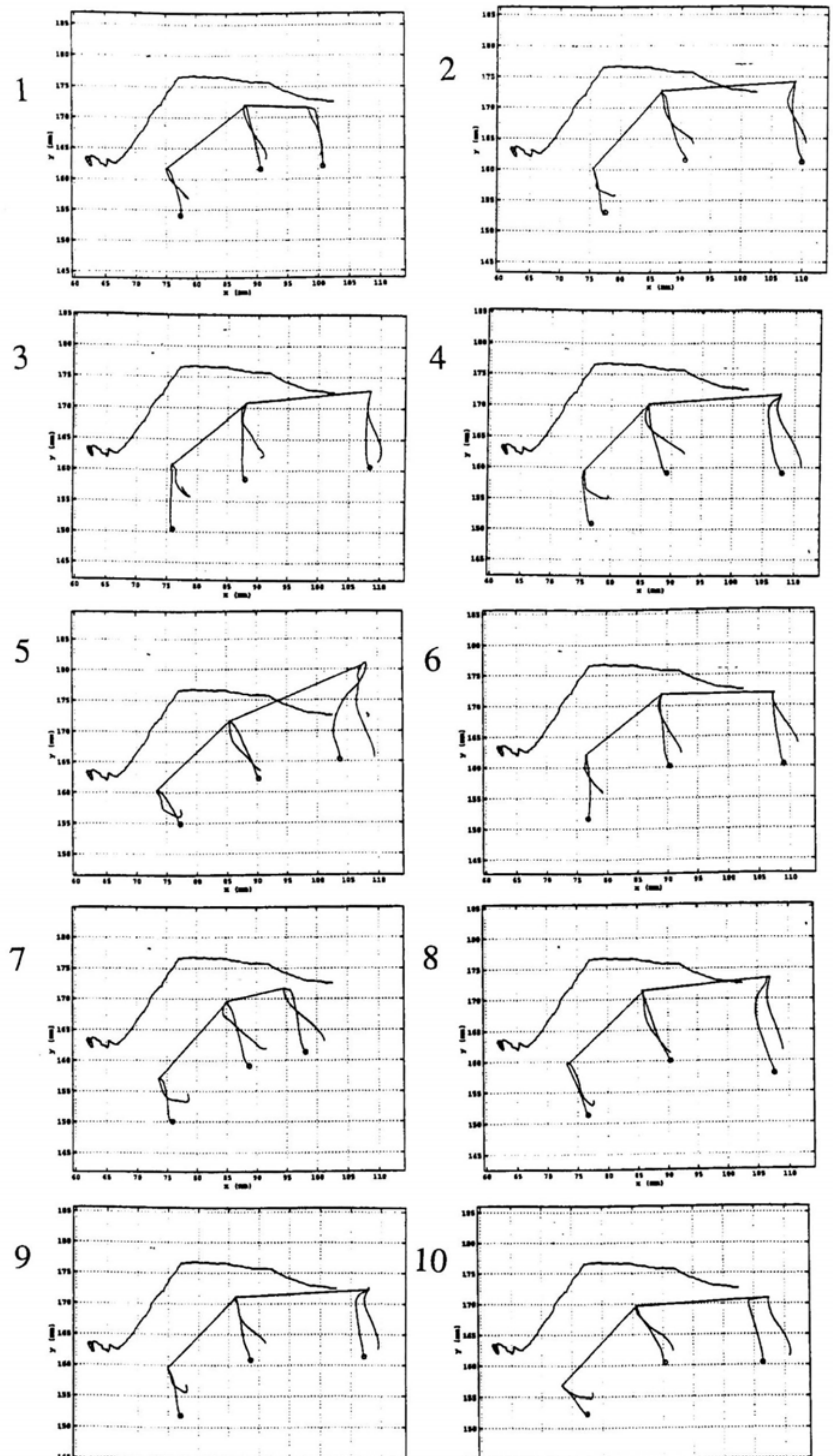


Figure 4.5 (i) Coil trajectory displays for individual fast speech control /ŋ#k/ tokens produced by ***subject D****. Movement paths of the tongue tip, tongue body and tongue dorsum coils are shown from the mid-point of pre experimental-sequence /a/ up to the mid-point of the following /ʌ/. The solid line is an interpolation of tongue configuration at maximum tongue dorsum displacement.*

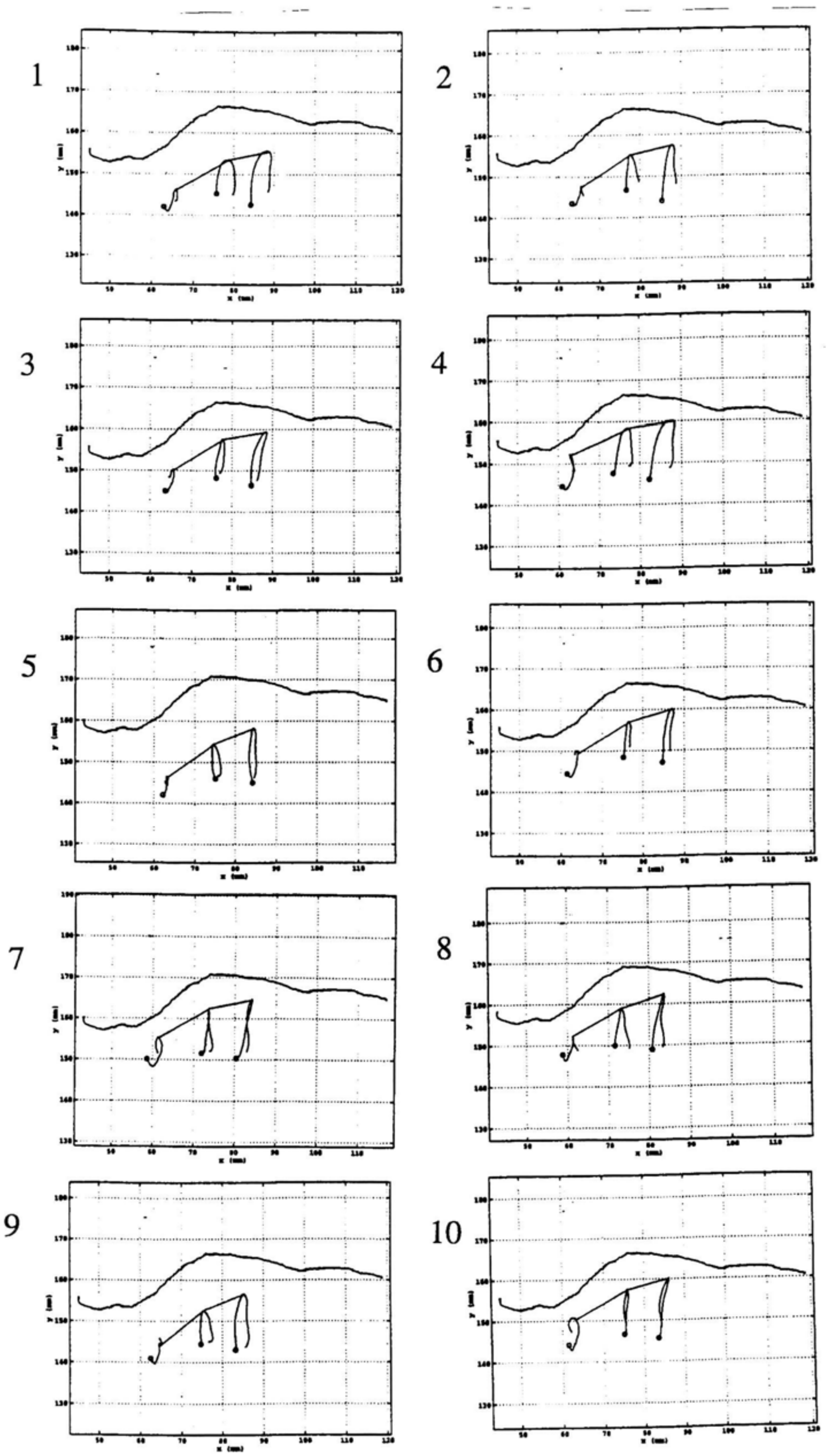


Figure 4.5 (ii) Coil trajectory displays for individual fast speech control /ŋ#k/ tokens produced by ***subject H****. Movement paths of the tongue tip, tongue body and tongue dorsum coils are shown from the mid-point of pre experimental-sequence /a/ up to the mid-point of the following /ʌ/. The solid line is an interpolation of tongue configuration at maximum tongue dorsum displacement.*

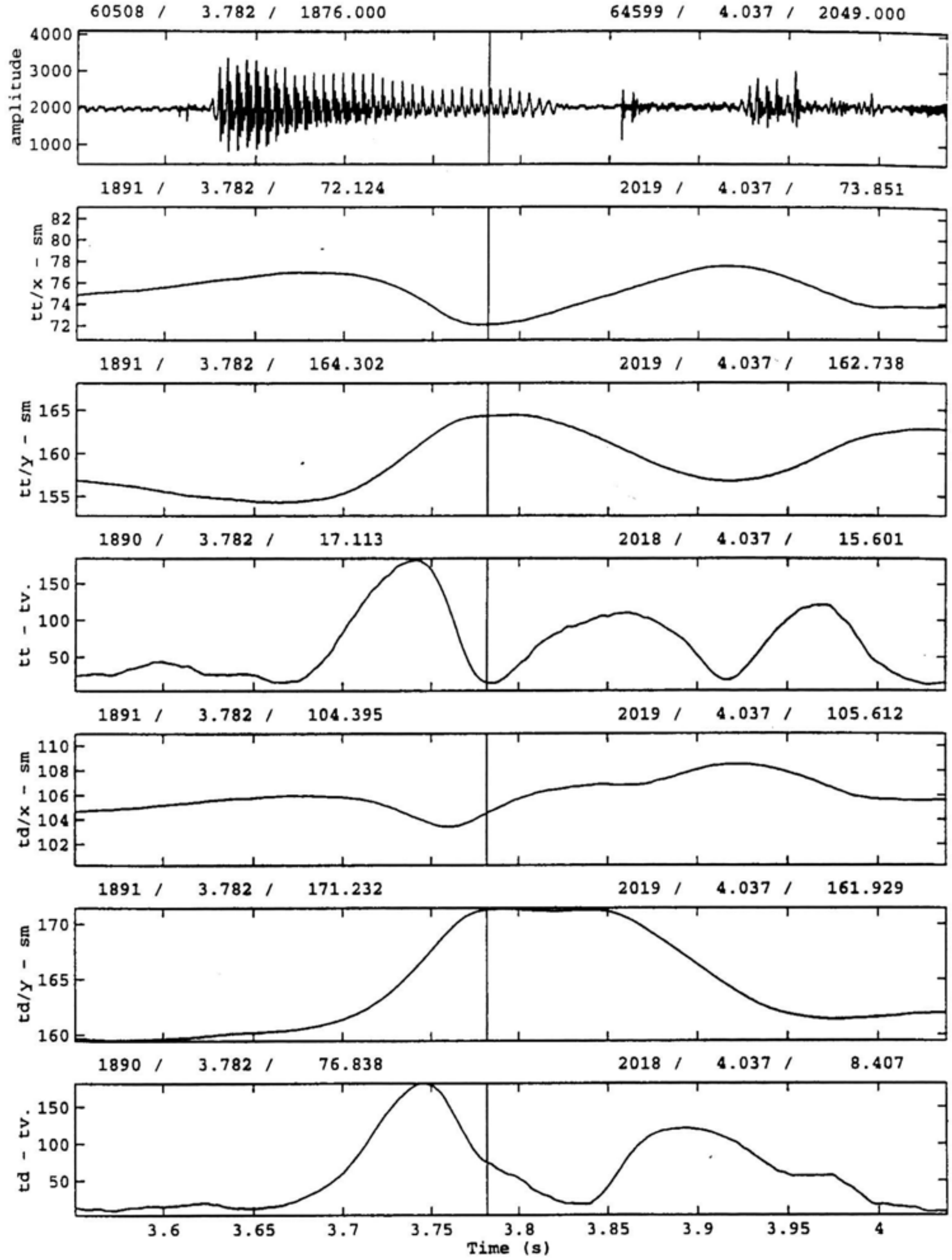


Figure 4.6 – waveform and EMA traces for one fast speech non-assimilated /n#k/ *token produced by subject D. Waveform at the top of the display and the EMA coil traces capture …*ban cuts…*The first two traces below the waveform show the tongue tip displacement trajectory in mm in the x axis (*'tt/x'*) and y-axis (*'tt/y'*) respectively. Below this (third trace down) is the tangential velocity trace (*'tt-tv'*) which shows combined velocity for movement in the x and y dimensions. The fourth and fifth traces below the waveform show tongue dorsum displacement trajectory in the x axis (*'td/x'*) and y-axis (*'td/y'*) respectively and the final trace shows the tangential velocity trace for the tongue dorsum (*'td-tv'*). All traces are time-aligned with the waveform. The solid vertical line across all panels indicates the moment of maximum tongue tip displacement in the x and y axis.*

For tokens where alveolar closure is achieved, (Subject D: numbers 1, 2, 3, 4 and 5 in Figure 4.4 (i)), it is at first sight difficult to discern any seriation of tongue tip and tongue dorsum gestures indicating that during a fast speech production of this sequence even when an alveolar stop followed by a velar stop is perceived, closure at the back and the front is achieved almost simultaneously. On closer inspection the tongue dorsum moves backwards horizontally after closure is made in order to consolidate the velar stop closure and to allow the tongue tip to drop so that velar release is audible. (This situation does not, however, apply to repetition 5.) Figure 4.6 illustrates this characteristic timing difference. At the top of the display is the waveform for the experimental sequence showing the vowel /a/, the closure phases, release and the following vowel /ʌ/. The first two traces below the waveform show the tongue tip displacement trajectory in the x and y-axis respectively. Below this (third trace down) shows the tangential velocity trace which shows combined velocity for movement in the x and y dimensions. The fourth and fifth traces below the waveform show tongue dorsum displacement trajectory in the x and y-axis respectively and the final trace shows the tangential velocity trace for the tongue dorsum. All traces are time-aligned with the waveform. The solid vertical line across all panels indicates the moment of maximum tongue tip displacement in the y axis (second trace below the waveform). We can see that at this point the tongue dorsum coil has also reached its *vertical* maximum (fifth trace below the waveform). However, the velocity trace for the tongue dorsum just below shows that the tongue dorsum is still moving in some dimension other than the y-axis and this further movement upwards corresponds to the horizontal backwards motion observed in the individual coil trajectory displays for the non-assimilations.

4.2.4 Timing measures

Since there is a timing difference between maximum distance travelled by the tongue tip and dorsum coil for 4 out of 5 of the non-assimilated tokens produced by subject D, it was necessary to see if such a timing relationship could be found for assimilated forms of /n#k/ as well. While the position graphs Figures 4.3 (i) and (ii) have shown no difference between assimilated /n/ and lexical /ŋ/ in terms of tongue tip position, it could be hypothesised that there is a subtle timing difference whereby, for the assimilated sequences, the tongue tip reaches its maximum displacement slightly before the tongue dorsum. The lexical /ŋ#k/ sequences on the other hand, would not show this timing. Table 4.1 below shows the timing of tongue tip maximum displacement (measured at minimum tangential velocity) relative to tongue dorsum maximum displacement for all tokens for both subjects. Each token is assigned a value in milliseconds which may be positive or negative, or is '='. The millisecond values indicate when in relation to the tongue dorsum the tongue tip has reached its furthest excursion for the experimental consonant sequence following /a/ for either /n#k/ or /ŋ#k/. For instance the first value for token 1 of /n#k/ for subject D is '56'. This means that for this unassimilated token the moment of tongue dorsum maximum excursion occurs 56 ms after tongue tip maximum displacement. The second value for token 2 of /n#k/ for subject H is '-22'. This means that tongue dorsum max. displacement occurs 22 ms *before* tongue tip max. displacement. The value for repetition 1 '=' indicates that the max. displacements occurred simultaneously.

Table 4.1 Time in ms between tongue tip maximum displacement (m.d.) and tongue dorsum m.d. for individual tokens of /n#k/ *and* /ŋ#k/ *fast speech produced by subjects D and H. '=' indicates that displacement maxima are simultaneous and negative values indicate that tongue tip m.d. occurs after tongue dorsum min. m.d. * = the only non-assimilations produced.*

subject D /n#k/		**subject D /ŋ#k/**		**subject H /n#k/**		**subject H /ŋ#k/**	
token	**value**	**token**	**value**	**token**	**value**	**token**	**value**
1*	56	1	12	1	=	1	-15
2*	53	2	=	2	-22	2	-17
3*	50	3	=	3	-23	3	=
4*	46	4	=	4	-68	4	-26
5*	23	5	=	5	-32	5	-16
6	=	6	=	6	-28	6	-24
7	=	7	=	7	-29	7	-19
8	12	8	=	8	-13	8	-14
9	=	9	13	9	-32	9	-40
10	=	10	22	10	-23	10	-35

The values for the first 5 tokens produced by subject D for /n#k/ (far left of table) are relatively high. For the purposes of this table and Figure 4.4 (i), this subject's non-assimilations are grouped together (tokens 1-5). This timing difference reflects the strategy described above for non-assimilated fast speech tokens. The assimilated tokens have either a very small value or no value at all, i.e. '='. The values for the /ŋ#k/ controls are likewise either very small or '='. The situation for subject H is very similar in that the assimilated /n#k/ and the control sequences show no timing differences (a t-test showed these to be non-significant). In fact the consistency of tongue configuration (indicated by the solid line in Figures 4.4 and 4.5) across the repetitions for both sequences is striking for this speaker. This rules out the possibility that the overall difference between derived and lexical /ŋ/ lies in the timing between front and back tongue movement.

There are some other observations that can be made from Table 4.1. The first is that the two speakers follow different trends regarding timing. For all tokens subject D either achieves maximum tongue dorsum displacement after maximum tongue tip displacement, or achieves these simultaneously, whereas for subject H, the situation is the other way around with maximum tongue tip displacement occurring after maximum tongue dorsum displacement or in two cases, these events are simultaneous.

4.2.5 /n#k/ careful speech assimilation

There was one /n#k/ assimilation in careful speech produced by subject H (repetition 10). This is shown in Figure 4.7 below.

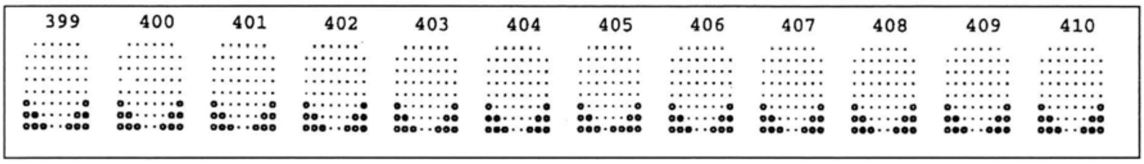


Figure 4.7 – EPG patterns for careful speech assimilation produced by subject H

Figure 4.8 below shows the tongue coil trajectory display for this token. The solid line represents the tongue configuration at the moment of maximum tongue dorsum displacement. Even though this is slow, careful speech, maximum tongue dorsum and tongue tip displacement are simultaneous.

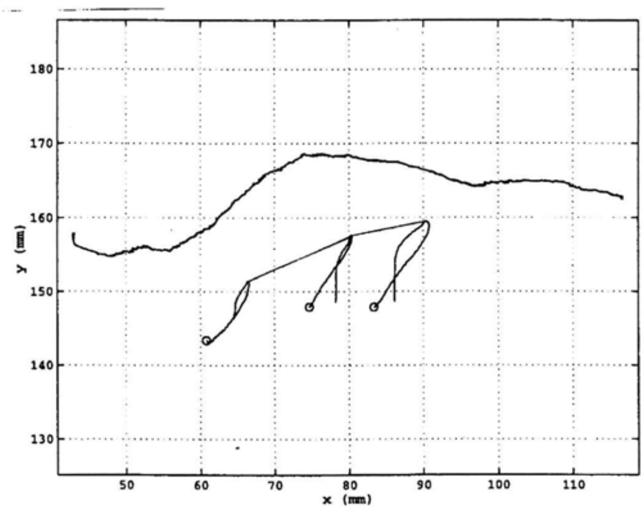


Figure 4.8 tongue coil trajectory for assimilated /n#k/ in careful speech produced by subject H.

In summary, the results of the follow-up EPG/EMA study showed that no residual coronal raising gestures for /n#k/ were found for either speaker. Somewhat surprisingly, for both speakers, the variability in tongue tip height at maximum displacement for control /ŋ#k/ was greater than for completely assimilated /n#k/. Furthermore, no differences in overall tongue configuration at the beginning of the experimental sequence between assimilated /n#k/ and /ŋ#k/ were found for either speaker. The lack of evidence of residual gestures in the EMA data has meant that judgements about individual-speaker strategies made from the original EPG-only study still stand. In the case of alveolar to velar assimilation, it appears that tongue-palate contact-only data is sufficient to reveal all assimilatory variants.

CHAPTER FIVE

DISCUSSION AND CONCLUSIONS

5.0 INTRODUCTION

This chapter presents a discussion of the results of both the EPG-only and the EPG/EMA follow-up experiment. The discussion will be structured around the Research Questions set out in Chapter 1 section 1.6. Discussion of Question 1 focuses on findings relating to inter and intra-speaker variability in production of the experimental sequence and also on the effect of speech rate on assimilation and on intergestural timing of the stop consonants for non-assimilations. With an understanding of speakers' 'performance', inferences can be made about corresponding internal processes, i.e. speakers' 'competence'. Discussion of Question 2 deals separately with implications for psycholinguistic modeling of speakers' assimilatory patterns and Question 3 assesses, in the light of the findings of this study, the adequacy of models of speech production which have particular relevance for place assimilation.

5.1 QUESTION 1:

How do speakers produce the sequence /n#k/?

Specifically in terms of:

- **effect of speech rate on spatial/timing description of the experimental sequence.**
- **amount of inter-speaker variation.**
- **amount of intra-speaker variation (individual speakers' assimilatory variation across repetitions).**
- **possibility of phonetic identity of underlying /n#k/ sequences and all-velar /ŋ#k/ control sequences.**

5.1.1 The effect of speech rate on the production of the alveolar to velar sequence and interactions with speaker-specific factors.

This section deals with the following: (i) speech rate as a factor in assimilation (the frequency and type of assimilations that occur in fast speech compared to slow/careful speech) (ii) speech rate as a factor in the intergestural timing of non-assimilations (iii) individual-speaker strategy preferences in terms of assimilation and in terms of intergestural timing (iv) the relationship between intergestural timing and assimilation.

In the EPG-only experiment, assimilation, whether partial or maximal, was completely absent in the slow/careful speech EPG data for 8 of the subjects. Only two subjects, G and H, assimilated in careful speech, subject G producing one and H producing three. These assimilations appeared to be categorical as there was no trace of the influence of the alveolar gesture in the EPG patterns. In the follow-up EPG/EMA experiment, subject H produced one complete assimilation in careful speech. No trace of an alveolar gesture was detectable either in the EPG or the EMA trace thus confirming that the assimilation was indeed complete. The assumption that this speaker's assimilations are complete in careful speech is extended to the assimilations produced by subject G in the EPG-only study. Therefore, in general, it was found that assimilation in careful speech is very infrequent and is categorical. These careful speech results conflict with Kerswill's findings (1985, discussed in section 1.3.2, Chapter One). In his EPG study of alveolar place assimilation he found that his single informant produced 5 residual alveolars and 5 complete assimilations out of a total of 20 tokens in 'slow and careful' speech. Thus assimilation in careful speech is both frequent and variable in type. However, Kerswill's stimuli consisted of nasal/oral and voiced/voiceless alveolar stops followed by either velar stops or bilabial stops. Also, the vowel context included the diphthongs /eɪ, əʊ/ and the vowels /ɜ, a/ and so the results of Kerswill's and the present study are not strictly comparable. Furthermore, comparable objective speech rate measures are lacking with which to compare the exact speech rate at which the careful speech was produced in Kerswill's study and the present one.

In fast speech, frequency of assimilation rises significantly from 4% of all experimental tokens in careful speech to 57% of all experimental tokens in fast speech (statistical analysis of speech rate in careful and fast speech confirmed that all speakers did increase their speech rate when instructed to do so). Furthermore, both complete and partial assimilatory forms were found in fast speech (although there were very few of the latter type overall). From an intra-speaker variability perspective, it was found that variability in /n#k/ forms within speakers increases in fast speech. Put simply, there are more speakers who vary between non-assimilation and assimilation in fast speech than in careful speech. In careful speech only two subjects, G and H, produce any spatial variation on alveolar stop, i.e. assimilation. In fast speech, however, four subjects, A, B, C and D, show variation in /n#k/ realisation. Of the remaining subjects, two do not vary from the production of non-assimilation (E and F) and four do not vary from the production of complete assimilations having opted to habitually re-organise the gestures in a radical fashion (G, H I and J).

These findings indicate that it is overall frequency of assimilation, intra-speaker variability and occurrence of assimilation types which distinguishes careful and fast speech. No residual alveolar articulations were found in careful speech which suggests that 'undershoot' forms of the alveolar stop can be firmly associated with a faster rate. This suggests that undershoot, or to use Lindblom's description 'duration dependent undershoot' (1963), comes about through unavoidable inertial effects on the tongue due to the demands of increased speech rate and not through a planned simplification. Some place this factor above all others (Hawkins, 1984).

The importance of the role of speech rate is also evident when comparing the speech rate (a continuous variable and not just the category of careful or fast) at which assimilations and non-assimilations were produced overall in fast speech. Figure 3.11, Chapter Three plots the speech rate for non-assimilations, residual alveolars and complete assimilations and a t-test showed that there was a highly significant difference between the rate at which fast speech non-assimilations were produced, speakers combined, compared with fast speech assimilations. This indicates the relationship between rate and achievement of the alveolar stop closure. This relationship is qualified by the fact that individual-speaker preferences can override this relationship.

One reason why a fairly strong relationship has been found between careful speech and fast speech is that two 'polarised' speech rate categories have been compared. In this sense the study has not presented a picture of 'normal' assimilatory behaviour (its occurrence in 'normal' speech rate). The design of the experiment may have inadvertently given the impression that careful speech data, with which the effects of increased speech rate can be compared, is somehow 'baseline'. But of course the slow/careful data are not representative of normal speech, not because the rate is slow (casual/conversational speech can be intermittently slow) but because the style is careful. If it can be assumed that speakers assimilate considerably more in natural conversational speech compared to slow/careful speech, then gesture economy (if this can be thought of as directly related to assimilation) must be a pervasive characteristic of normal speech. The idea that the principle of gesture economy underlies speech production, has been most notably argued by Lindblom (1983). Barry (1985) and Kerswill's (1985) EPG work on alveolar to velar assimilation compared the frequency and type of assimilation produced under various speech rate conditions. The results are reproduced again here below (summary taken from Nolan, 1992: 268). On the left is shown an average of 3 speakers who produced the stimuli at 'normal' and 'fast' rates (Barry, 1985) and on the right is shown articulation types for a single speaker speaking slowly and carefully, normally, fast and carefully and fast (Kerswill, 1985). Barry and Kerswill used the same definition of alveolar stop closure as the present study uses.

Table 5.1 EPG articulation types at 2 different rates from 3 speakers combined (Barry, 1985) and on the right, Kerswill's data (1985) showing articulation types for 1 speaker at 4 different rate/styles (from Nolan (1982: 268)

	Normal	*Fast*	*Slow & careful*	*Normal*	*Fast & careful*	*Fast*
Full alveolar	23	15	10	2	3	0
Residual alveolar	14	15	5	8	3	2
Zero alveolar	11	18	5	10	14	18

These data show that while for all speakers there are more assimilations (residual alveolars and complete assimilations) at fast rates, assimilation is common at normal rates. For the single speaker on the right-hand side of Table 5.1 there are only 2 more assimilations (partial and complete) produced at the 'fast' rate than there are at the 'normal' rate. The approach taken in the present study is that the slow and careful condition should be considered a special case in relation to assimilation, where an 'override' of normal assimilatory behaviour is required in the form of full alveolar stop production. Incidentally, the criteria for identifying tokens as residual alveolar types from EPG data used by Barry (1985) and Kerswill (1985) was also used in the present study (complete alveolar closure is absent but there are more EPG contacts in the direction of the alveolar ridge than for the control velar to velar tokens).

The discussion so far has focused on identifying overall trends that are associated with differing speech rates but the results of this study have thrown up important effects that can only be attributed to individual-speaker preferences. The fact that subjects G and H assimilated in careful speech shows that assimilation is not motivated by speech rate alone for all speakers. If assimilation can occur when gesture economy is not so much of a priority, as it is presumed to be in careful speech, then phonetic context alone can be enough to trigger assimilation. Importantly, these two subjects always assimilated in the fast speech condition indicating that assimilation is a more pervasive feature of these subjects' speech compared to others and that they have a lower 'threshold' below which speech rate is too slow or careful to trigger assimilation. Other important evidence that speaker preferences can override overall speech rate trends, is that subjects E and F never opt to assimilate in fast speech. In these cases it is clear that a measurably faster speech rate does not automatically result in a tendency to assimilate. Figure 3.11 in Chapter Three shows that subjects E (a non-assimilator) and I (a 100% assimilator) have tokens in almost identical speech rate ranges (statistically confirmed) and yet one exclusively applied a complete assimilation strategy and the other exclusively applied a non-assimilation strategy.

Discussion will now focus on production of non-assimilations. In a previous study on the effect of speech rate on the production of a variety of stop consonant sequences Byrd and Tan (1996) found that speech rate had little effect on articulatory overlap in the production of non-assimilated /d#g/ clusters. They commented: 'this presumably is due to a ceiling effect, whereby /d#g/ is so overlapped at slower rates that only minimal additional increase due to faster rate is evidenced.' (p.270) In the present study, only 46% of all careful speech non-assimilated /n#k/ tokens showed any overlap of the stop closure phases (that is, a period of simultaneous front and back EPG closure). However, it is not possible to compare these results properly since Byrd and Tan's criteria for overlap includes part of the build-up towards the closure phase for C_2. Overlap in their study corresponds to a spatial measurement: 'the percent of C_1 contact [in a C_1#C_2 cluster] during which C_2 contact occurred' whereas in the present study overlap corresponds to the presence and duration of simultaneous alveolar closure and maximum velar constriction. Even though the components of overlap in the present study were targets, not including movement towards that target, it is still worth comparing the incidence of overlap types in relation to other non-assimilation types. There were slightly more 'simultaneous' tokens (50%) in careful speech compared to overlap tokens overall (the former involves simultaneous alveolar release and velar closure) but there were very few (4%) serially-ordered sequences indicating that the trend is strongly against this type of timing organisation for careful speech. As for the effect of increased speech rate, the results of the present study indicate that the percentage of all non-assimilations showing EPG-defined overlap decreased from 46% in careful speech to 35% in fast speech. It must be noted here that since the identification of the closure and release of each consonant in the sequence is based on EPG tongue–contact data alone it is likely that a substantial part of the velar gesture (i.e. the movement towards the target) has already taken place before actual full contact for the closure has been made. Another reason why these studies are not strictly comparable is that Byrd and Tan were looking for changes between 'normal' rate and 'fast' rate of speech whereas in the present study the slowest rate was slow and careful which possibly explains the comparatively low frequency of overlapped sequences. In general, speakers varied considerably between the three non-assimilation types. The variation most noticeably occurred between speakers indicating that individual speaker-preferences are apparent for production strategy on the level of assimilation and on the more fine-grained level of contiguous lingual stop closure timing.

The possibility of a relationship between intra-speaker variability on this level (the timing relationship between stop closure and release for the two adjacent consonants) and intra-speaker variability in assimilatory forms in fast speech has not been clearly supported by the results of this study. The comparison is made because if there was a correlation between these levels for those speakers who showed more variability in assimilatory forms (i.e. subjects A and B the 'gradual assimilators') than others then the interpretation could be made that variability in assimilatory forms is a corollary of a relatively high level of speech production variability. Table 3.11, section 3.2.1, shows that subject A produced the biggest spread of careful speech non-assimilation types of /n#k/. This speaker also produced a full range of assimilatory forms in fast speech including partial assimilations. However, this is not the case for subject B who stuck to the 'simultaneous' type for all but one careful speech repetition but who produced a similarly full range of assimilatory forms in fast speech. A question arises here as to how meaningful these comparisons can be since the non-assimilation types are qualitative abstractions taken from a timing continuum, in the sense that a serial-ordering strategy is only gradually different from a 'simultaneous' strategy and that an overlap strategy is only gradually different from a 'simultaneous' strategy. The only difference is in timing of velar closure in relation to alveolar closure. It is unlikely that these EPG-defined variations would be discerned in the corresponding EMA data. This discussion also leaves aside the question of the perceptual relevance of these variations.

The other parameters which speech rate can influence are the presence and duration of double articulation and the temporal latency (cl1-cl2) for non-assimilated tokens of /n#k/. A Spearman rank correlation found that there was no relationship between the continuous variables of speech rate and duration of double articulation in careful speech tokens or fast speech tokens. As for temporal latency, no correlation was found for the careful speech tokens overall between speech rate (continuous variable) and cl1-cl2 duration such that as rate increased the time it took for velar closure to be achieved following alveolar closure decreased. No such correlation was found either for the fast speech tokens nor were correlations found at the level of individual speakers in either speech rate sample. It also turned out that this decrease in cl1-cl2 was not associated with assimilation. It appears that the process of assimilation is unrelated to quantitative changes in this phonetic parameter associated with coordination of the experimental sequence.

5.1.2 Articulatory details of /n#k/ assimilation and variability

Previous instrumental studies in this area have tended to give only cursory attention to variability at the level of the individual speaker but this study shows that it is of primary importance with significant ramifications for internal modeling of assimilation. Previous EPG studies as far back as the late 1970s (Hardcastle and Roach, 1979) and as recently as 1996 (Kühnert, 1996) have noted partially assimilated alveolar stop closure targets. These observations were usually made in the context of showing that the articulation of alveolar to velar sequences does not mirror the discrete phonological modeling of the assimilation process. The assumption underlying these studies, however, since it is not stated otherwise, is that assimilation, where it occurs, is a gradual process for *all* speakers. Indeed, Kühnert states on the basis of her alveolar to velar EMA data (and not on the basis of her EPG data which was simultaneously collected with the EMA data) that 'all speakers [numbering 5] execute residual coronal movements at least some of the time' a fact that she considers supportive of a gestural blending account of assimilation. These studies, in particular the EPG studies, were primarily concerned with either the effect of speech rate on frequency of assimilation or whether the frequency of assimilation depends on the alveolar stop being voiced or nasal or part of a stop sequence. Neither of these aims required specific focus on the level of the individual speaker although speaker-specific preferences for frequency of assimilation had been noted. Holst and Nolan (1995), however, challenged the prevailing idea (which Nolan had originally contributed to) that all connected speech processes are the result of gestural overlap and are thus phonetic processes. They proposed, from acoustic evidence of /s/ to /ʃ/ assimilation, that across a group of speakers assimilation is mainly due to the coalescence of the gestures (the kind of mechanical dynamic output of a device required to reach two incompatible targets in a short time) but that it can sometimes be the result of a cognitive rule leading to /ʃ#ʃ/. Nolan (1986:7) proposed that reduction can be gradual until a threshold is reached where 'phonological restructuring intervenes to modify phonological forms (either at the level of underlying representations, or within the phonological derivation).' This mixed approach 'fits in with a view of phonetics which embraces variation in production, perceptual reinterpretation , and sound change.' (p.330). It conveniently dispenses with the need to explain exactly why different productions of the *same sequence* produced by the *same speaker* (since Holst and Nolan do not look for or report any speaker-specific effects) can be either partially assimilated or completely assimilated, with the latter forms requiring a completely separate strategy. The idea that a phonological rule is triggered after a certain level of reduction is reached is not supported by the findings of this study. Alveolar assimilation is not caused by extreme reduction of the phonetic parameter of temporal latency (interval between alveolar closure and velar closure). The present study has taken a closer look at intra-speaker assimilation to discover that it is not possible to put forward a unifying account of alveolar to velar place assimilation as either a phonetically motivated process or a phonological one.

A general observation made from this study is that there is a contrast between speakers who show some variation in alveolar realisation within their 10 fast speech /n#k/ repetitions and those speakers who show no variation. But, the findings also show that there is a much more intriguing contrast on a finer level. Subjects A and B (gradual assimilators) produce a range of forms including non-assimilations, partial assimilations and complete assimilations while subjects C and D (binary-option assimilators) alternate in a binary fashion between non-assimilation and complete assimilation. Assimilatory strategies to increase speech rate in the production of /n#k/ can take the form of a categorical restructuring of specified gestures or a spatial diminution of the alveolar gesture, but the important point is that speakers appear to implement either one or the other of these strategies. Unlike Holst and Nolan (1996), who do not specify how the gesturally blended or categorically assimilated forms of /s#ʃ/ are distributed across and within speakers, this study shows that these strategies do not compete within speakers. This finding and the implications it has for psycholinguistic modeling will form the central discussion issue for this dissertation.

5.1.2.1 The 'non-assimilators'

The fact that subjects E and F never assimilated indicates that assimilation is not motivated by speech rate for all speakers. Table 3.11 in section 3.2.1 showed that the trend for these speakers is to produce more 'simultaneous' tokens (where velar closure is achieved as the alveolar closure is released) in fast speech. This shows that overlap does not simply become more extensive as the speech rate increases. Instead it appears that there is a tendency to finely coordinate the offset of one target with the onset of another. If we also look at the durations of alveolar and velar double articulations expressed as a proportion of the interval between the beginning of /a/ in *ban* up to the end of /s/ in *cuts* shown in Figure 3.22, we can see firstly that apart from one outlier value where the percentage double articulation is as much as 35%, subject E's variability range for percentage double articulation of the larger utterance is more or less identical for careful and fast speech.

5.1.2.2 The '100% categorical assimilators'

These speakers each produced apparent completely assimilated /n#k/ sequences for all 10 of their fast speech repetitions. Assimilation for these speakers appears to be a categorical and habitual process with no discernible trace of an alveolar gesture left in the contact patterns. In the follow up EPG/EMA study, no residual alveolar gestures were found for subject H. The consistency of these /n#k/ repetitions produced in the EPG-only experiment in terms of duration of preceding vowel /a/ and duration of velar closure phase can be seen on the timing bars for these speakers on Figure 3.16 (vi)-(x). These discrete assimilations appear to bear out feature geometry modelling of assimilation thus :

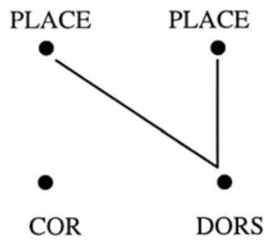


Figure 5.1 - autosegmental representation of alveolar to velar assimilation

Here, the place node assigned to the coronal consonant has been delinked and has been associated instead with the dorsal consonant, due to the 'spreading' of the place feature for the dorsal consonant rightwards. The concept of 'spreading' is powerful in phonology but offers no insights into why this phenomenon might occur.

It seems very likely from the articulatory data that the habitual short-cut adopted by the 100% categorical assimilators is the result of a high order phonological rule. There is no evidence that any motor commands associated with /n/ are sent to the articulators. This result also means that the judgments made on the basis of EPG patterns alone (i.e. tongue-palate contact only) in these experiments are reliable. In this case, no new insights have been gained from the extra kinematic information beyond that which tongue-palate contact data can provide at least as far as presence or absence of residual alveolar gestures is concerned.

The hypothesis that differences in lexical form will always result in differences in phonetic form (Nolan, 1992, although he has since distanced himself from this notion) cannot be rejected or accepted on the basis of comparisons made in this study between the phonetic properties of derived and lexical /ŋ/. For subjects H and I statistical testing of the durations showed a significant result in the direction of lexical /ŋ/ being longer than derived /ŋ/ (see Table 3.7). However, there was no significant difference for subjects G and J. It is interesting to note here that the difference for H and I is in the opposite direction to that found by Holst and Nolan (1995) for /s#ʃ/ assimilation. They found that assimilated cases (/ʃ#ʃ/) were longer in duration than underlying /ʃ#ʃ/ (see section 1.3.5). Wright and Kerswill (1989) have suggested that a phonetic 'residue' of the alveolar is not only present but can be utilised by listeners to recover the alveolar in the case of assimilated alveolar to velar sequences. A comparison of individual-speaker summaries of EPG contact for derived /ŋ/ and lexical /ŋ/ over 10 repetitions (Figures 3.6 (i) and (ii), section 3.1.2.1.1) showed that there were no discernable place-of-articulation differences between the two forms. However, on close inspection of these prototypical frames, three of the four 100% categorical assimilators (H, I and J) who always completely assimilated in fast speech showed that for some repetitions of /n#k/ the velar pattern was slightly more retracted than for /ŋ#k/.

While the data reduction summaries of raw EPG data shown in Figures 3.6 (i) and (ii) are not available for EPG data acquired through the combined EPG/EMA methodology, an assessment can still be made on whether retracted velar patterns for /n#k/ occurred for subject H and, more importantly, if in fact this was due to a high tongue tip position. As it turned out there were no retracted patterns. There was no justification, then, for pursuing a correlation between tongue tip height and position of velar closure for any of subject H's tokens. Unfortunately, we can do no more than speculate about this relationship for the velar retraction found for /n#k/ in the EPG-only study for subject H, I and J.

Individual token displays of tongue tip, body and dorsum trajectories from the EPG/EMA study revealed a striking consistency in subject H's productions of both fully assimilated experimental sequences and control sequences. Table 4.1 showed that the tongue tip only reached its maximum displacement once the tongue dorsum had done so in all but two tokens (where the timing of this was simultaneous). For one token the time lag between these two is as much as 68ms (/n#k/ repetition 4). It is interesting to note that the results for Subject D's timing of these events show that tongue tip maximum displacement occurs slightly *before* the tongue dorsum has reached its maximum. The highest values for subject D occur for the /n#k/ non-assimilations where this timing relationship is most likely. It is difficult to speculate on why the two speakers have contrasting timing trends and it would certainly be difficult to relate these results to the properties of residual alveolars. It is possible that this reflects anatomical/biomechanical differences between individual speakers.

The individual-token trajectory displays have also allowed an assessment of whether speakers show a raised tongue tip and a hollow tongue body configuration. At no point in any assimilated /n#k/ token for either subject D or H was the tongue body coil lower than that of the tongue tip and so this type of residual articulation was not found. It seems that this hypothesised tongue configuration is not in fact borne out in the data and is maybe an unrealistic expectation anyway.

An interesting and unexpected effect shows up on the tongue tip and tongue dorsum position graphs for subjects D and H (Figures 4.3 (i) and (ii)). For both speakers it appears that the x-y position spread for the tongue tip for assimilated /n#k/ tokens (green triangles) is on the whole 'tighter' and less variable than the position spread for the lexical /ŋ#k/ tokens (blue triangles). For subject D the assimilated tongue tip positions are more constrained vertically than the equivalent positions for the underlying velar tokens while for subject H the effect seems to be stronger. This trend for both speakers brings about an interesting situation where the tongue tip for the assimilated alveolar cases seems to behave more as if there were a target to be achieved, that of no displacement.

5.1.2.3 The 'gradual assimilators'

Assimilation for subjects A and B is a gradual, phonetic process. Their EPG patterns for /n#k/ in fast speech can be ranged along a continuum from full alveolar closure to complete assimilation involving velar closure only. In between there are intermediate assimilatory forms where the tongue's target for /n/ has been undershot and is thus only partially executed. The range of /n#k/ forms was found to be unrelated to variability in speech rate.

It is clear that for the gradual assimilators the process of assimilation does not involve the implementation of a high-order phonemic substitution rule and thus a radical restructuring (i.e. simplification) of commands to the vocal tract. We can surmise from this that the alveolar 'specification' persists into the phonetic plan for these speakers, although at some level it undergoes various degrees of reduction in displacement. A question arises here, however, regarding the nature of the tokens at the far end of their assimilation continuum. These tokens showing velar stop patterns indistinguishable from lexical velar stop patterns at first suggest a categorical complete assimilation. But taken in context these tokens can be interpreted simply as extreme forms at one end of a continuum of gradual reduction. This dispenses with the requirement to invoke a separate phonological rule in addition to gradual alveolar weakening to account for all intra-speaker allophones of /n/ for subjects A and B. If this were the case, it would be an unparsimonious approach to describe *within*-speaker forms of /n#k/ but the evidence from the present study unambiguously points to the existence of two separate 'origins' for assimilation depending on speaker.

5.1.2.4 The 'binary-option assimilators'

The binary-option assimilators make up the other half of the subjects who sometimes assimilated and sometimes did not in the EPG-only experiment. While assimilation in fast speech for subjects A and B is gradient, subjects C and D are clearly less tolerant of /n/ allophones and demonstrate only categorical assimilations. As is the case for the 100% assimilators, C and D's assimilations are explained as the application of a cognitive rule. While both the gradual assimilators and the binary-option assimilators produce complete assimilations, which appear to be the same on EPG records, they are viewed as arising from fundamentally different strategies. The binary-option speakers' complete assimilations have to be considered the result of a high-level rule because, unlike the gradual assimilators, these assimilations are not at the end of a continuum of gestural reduction. Therefore the identification of partial assimilations in between non-assimilation and complete assimilation, is a crucial factor. A challenging issue arises when consideration is given to how the application of this optional cognitive rule is triggered. This question does not arise in the same way for the 100% categorical assimilators because the alveolar to velar assimilation rule appears to be 'switched on' by default for these speakers in fast speech.

The implementation of this categorical assimilation strategy in fast speech was confirmed by the EPG/EMA data for subject D. There was no evidence of tongue tip raising for the assimilations produced beyond that for the velar controls and so there were no intermediate partially assimilated alveolar productions. For the non-assimilations, the whole tongue appears to move as a single articulator at the beginning of the consonant sequence. This observation supports Marchal's hypothesis (1988) that stop sequences of the type VC_1C_2V are not produced in a linear order but as a cohesive unit. This hypothesis opposes the basic theory of coarticulation which assumes that the input to the speech production system is discrete and invariant units, the boundaries of which are eventually obscured during production itself. Instead, Marchal proposes that:

> The consonantal mechanism initially produces a rather neutral consonantal gesture: the whole lingual body moves upward, and one cannot assess at first whether or not it is preparing for the articulation of C_1 or C_2. The tongue first seems to adopt a stable position where it can presumably obtain tactile feedback about its current location and articulatory target. It can then differentiate the upcoming articulations, and produces C_1. (p.293)

5.1.3 Methodological considerations

This section discusses: identification of residual alveolars from tongue-palate contact data only and from EMA data and the inferences that can be made about tongue configuration; identification of residual alveolars; whether these two types of incomplete assimilation are essentially the same and reasons why these forms are detected differently and finally, the significance of these observations for modeling assimilation.

The definition of a residual alveolar gesture on the basis of individual 2-dimensional tongue trajectory displays and thus the confidence with which we can reject vertical tongue tip displacement accompanying assimilated /n#k/ tokens as 'non-linguistic', is one that Kühnert (1993) has pursued in a very different way. For her, the motivation for scrutinising tongue trajectories for alveolar to velar sequences was to establish whether two identical EPG patterns, for instance, one an assimilated realisation of /t#k/ the other a realisation of /k#k/, could have a quite different underlying tongue configuration. The purpose of this was to answer the question of whether different lexical forms always result in different phonetic forms. If Kühnert found that the tongue configuration for assimilated tokens were always different from control velar to velar sequences e.g. /k#k/ but that the two lexically different forms had indistinguishable velar EPG patterns, then this hypothesis would be valid. However, Kühnert found that this was not always but only *sometimes* the case. This is a valid judgement on the basis of comparisons between tokens. By contrast, the most important concern of the present study is to define a alveolar gesture in relation to normal variability of tongue tip movement for control sequences.

On the subject of taking EPG-defined assimilated tokens and looking again for evidence of residual alveolars in the EMA data, doubts must be raised about the validity of this approach. There are three factors which need to be considered. Firstly, the data from the present study (see Figures 4.3 (i) and (ii)) have shown that for lexical velar-velar tokens, the tongue tip position when the back of the tongue has reached its furthest excursion is considerably variable vertically and for both subjects shows the highest positions in comparison to the tongue position for assimilated /n#k/. This suggests that the tongue tip, when not executing an actual stop closure, is not constrained. It seems, then, rather arbitrary to consider that any tongue tip positions for assimilated tokens which only slightly higher than the lexical tokens should then be considered evidence of incomplete assimilation. This is the approach adopted by Kühnert (1993) and the one adopted in this study. Because lexical /ŋ/ tokens showed higher tongue tip positions than the assimilated /n/ tokens in the present study, the situation never arose whereby any residuals could be identified. A second reason for doubting the measurement of tongue tip height as an indicator of a residual alveolar, is that there is no control for the influence of speakers' jaw position at the moment of articulating this sequence. That is, if a speaker has a more open jaw position then the highest point of a residual alveolar gesture may not be as close to the alveolar ridge as other tokens. Perhaps it is distance travelled by the tongue tip that is a more significant measure than final position and perhaps it is a more closed jaw setting which is manifested as EPG-defined residual alveolars. This issue does, however, raise questions about the complex dynamic interactions between the jaw and the tongue, which is a subject beyond the scope of this dissertation. Thirdly and lastly, it has been shown that in the production of a non-assimilated /n#k/ sequence, the tongue initially raises as a whole 'unit' and differentiation of the front and back only occurs once the tongue is or is almost making contact with the hard palate. On the basis of this it is difficult to see how, for the production of the type of residual articulation not captured by EPG, the tongue tip/blade gesture, independent of the velar gesture, might stop short of its target. It would be easier for this gesture to be 'cancelled' at the outset as the application of a phonological rule would ensure.

5.1.3.1 Types of residual alveolar

At this point in the discussion it is appropriate to review what articulatory forms residual alveolar articulations can take. The first type, identified from EPG which pre-dates the use of movement tracking devices, are intermediate forms of alveolar stop closure where the tongue is high, while the tongue tip/blade fails to make median closure on the alveolar ridge. It was very surprising to find that there were only 2 such articulations identified from the present study. The other type of residual articulation observable with EMA used in conjunction with EPG (the latter is needed to determine whether closure has occurred or not) may involve a different tongue configuration. Here the tongue tip raising gesture is undershot and so displacement displays would show an underlying high vertical position for it while the EPG pattern shows only a velar stop pattern. Since the EPG pattern will not have shown any evidence of this gesture a rather hollow tongue body is anticipated unlike the EPG-defined residual cases. Thus the two forms would contrast in terms of how 'front' the tongue body is.

That residual forms are identifiable from EPG contact patterns is well attested by previous EPG studies (see section 1.3 for a review). But while the idea of the second type of residual articulation seems fairly established, convincing evidence of it is less apparent. The EMA-defined residual inolving a 'hollow' tongue configuration is probably very rare. Perhaps a more profitable way of looking at this question is to view the two types as essentially the same in that both involve a high tongue tip just short of closure. The factor which may differentiate the two reduced forms is the anatomy of individual speakers' vocal tracts. It may be that a flatter and possibly broader palate may give rise to the type of residual alveolar observed from EPG patterns simply because as an alveolar gesture is initiated there is more opportunity for the sides of the tongue to come into contact with the hard palate. It may be that more sophisticated speech imaging techniques still are needed to get to the bottom of this question.

In section 1.3.3 (Chapter One) Kühnert's assertion that assimilation is for all speakers gradual was questioned. From the articulatory position displays of hers shown in Figure 1.14, it was argued that speaker GER3 is not 'gradual' in the same way as GER1 is, as had Kühnert claimed, because for GER3 there was significantly more overlap between the position ellipsis for the velar controls and the position ellipsis for the assimilated alveolar tokens. Kühnert does, however, explicitly state that the displays for the two speakers are different and that the vast majority of the data for GER3 seems to support absolute gestural disappearance. In the light of the results of the present study it could be argued that GER3 is in fact applying a phonological gestural deletion/substitution rule while assimilation is indeed for subject GER1 gradual.

A drawback of Kühnert's study is that the full set of EPG patterns are not shown from which it may have been possible to observe EPG-defined residual alveolars. Another reason why inclusion of the EPG patterns in Kühnert's work would be useful is that it would be interesting to discover whether contact patterns for alveolar to velar sequences such as /d#k/ surrounded by non-low vowels show residual alveolar articulations on the EPG patterns themselves in view of the fact that she proposed that low-vowel environments seemed to produce the type of residual alveolars that did not show up on the contact data. In her experiment stimuli included all combinations of the following vowels: for the German subjects V_1 = /a, ɛ, ɪ, ɔ, ʊ/ and V_2 =/ɑ:, a, ʊ/ and for the English subjects V_1 =/ʌ, e, ɪ, ɒ, ʊ/ and V_2 =/ɑ:, ʌ, ʊ/. Thus it would be possible to compare, say, the production of a sequence like /ɪt#kʊ/ with a sequence like /at#kɔ/.

5.2 QUESTION 2: At what level in the generation and execution of an utterance does assimilation occur?

The existence of speaker-specific assimilation strategies is not a new observation and inter-speaker coarticulation strategies have also been noted (see section 1.4 of Chapter One) but what makes the finding that some speakers assimilate gradually and some categorically so significant are the implications it has for *where* in the 'speech chain' (from the access of abstract lexical representations in the planning stage of an utterance to their realisation as physical speech movements) place assimilation is located. Moreover, there are implications for how alveolars, as phonetically and phonologically special items, are cognitively represented, particularly when they appear in the context of a following velar stop. What follows is a discussion of these issues and a proposal that two separate planning processes should be invoked to account for the two opposing assimilation strategies found. It will be shown that the answer to the above question 'At what level in the generation and execution of an utterance does assimilation occur?' is that it depends on the speaker.

The most appropriate way to view the data from the gradual assimilators is to consider it a natural consequence of the articulators rapidly achieving two contiguous articulatory targets with the same organ, under conditions of increased speech rate, and thus phonetic weakening takes place at the physical execution stage itself. These speakers' underlying representation of the alveolar to velar sequence has not undergone radical restructuring and the alveolar target is retained. We can infer from this that the 'intention' to produce it is always present but a weakening rule allows reduction of displacement when the gesture is executed. Since English is a language tolerant of allophonic variation for alveolars, it is this item in the sequence which is truncated in order to produce it more quickly. Current understanding assumes that this weakening rule is located further towards the intiation of speech motor commands than a higher-order phonological rule. This level is presumably where other language-specific effects such as devoicing of oral stops and de-aspiration of voiceless oral stops in clusters such as *spin* are located.

Whereas assimilatory variants can be thought of as 'computed on-line' for the gradual assimilators, for the binary-option assimilators they are 'stored' alternatives. The optional categorical assimilations produced by the latter group must be generated at the abstract planning stage of speech production. This formalism allows no trace of an alveolar gesture to surface in the phonetics and this process has taken place before these abstract values are formed into a phonetic plan. This segment-sized velar allophone is presumably stored along with other variants on the alveolar to be accessed depending on phonetic context (e.g. /n/ becomes [ɱ] in *information*). The notion of stored allophonic alternatives is reminiscent of Wickelgren's ideas (1969). He proposed that speakers have an inventory of stored context-sensitive allophones rather than only phoneme-like segments. These stored allophones are variants of phonemes which come about through potential left and right context effects. As an explanatory framework for speech production in general this theory soon falls down. It does not address coarticulatory influences beyond neighbouring segments and would involve huge numbers of elements to be stored. However, when it comes to assimilation (of the complete variety seen in the present study) the idea of stored allophones appears to make sense.

While the composition and cognitive storage of these allophonic substitutes of /n/ are difficult concepts, even more difficult is the question of how the rule which retrieves these optional alternatives is triggered. A way around this problem would be to suggest that the speech production system is defaulted to plan for complete alveolar assimilation following the principal of phonological underspecification. In this way it would be the production of non-assimilations that required an override operation. The advantage of this idea is that it complements intuitions about speakers' level of awareness when producing careful and casual forms. It is traditionally held that alveolar to velar assimilation is not a conscious process for speakers and listeners (although this is not the case for all connected speech processes, for example the production of a glottal stop in place of a voiceless alveolar plosive in some contexts acts as a sociolinguistic marker and therefore is highly salient) whereas the production of unassimilated alveolars is often 'conscious', being associated with intentional enhanced intelligibility.

The view that assimilation is 'phonological' for some speakers and 'phonetic' for others (for want of better descriptors) contradicts the central premise of Articulatory Phonology that all possible assimilatory forms stem from a single source, i.e. gestural blending. Coarticulatory and assimilatory forms originate from the nature of speech production itself, namely its intrinsic organisation of time. But the findings in this study are not satisfied either by standard phonological formalism. A likely explanation is that the gradual assimilators either lack the application of a cognitive operation that handles the optional substitution of one segment for another or 'override' it by another operation (alveolar weakening) at a lower processing level. It is tempting, however, to consider that the issue is one of a deficit in stored allophones if the view is taken that an on-line phonetic alveolar weakening rule is in place anyway for all speakers of English, at least.

Hayes (1992) in his response as a phonologist to Nolan's gradient alveolar to velar EPG assimilation data (1992), discussed in section 1.3.2 (Chapter One), clearly separates phonetic rules from phonological rules in order to reach what he considers to be a solution. Nolan had rejected phonological accounts of place assimilation (more specifically, complex segments where the place features for the coronal are linked to those of the dorsal without losing the coronal features altogether) because (a) they fail to capture varying degrees of accomplishment of the alveolar target and (b) they cannot show why it is the alveolar stop that can be weakened while the following velar is immune to reduction. Hayes replies by saying that the phonological level is not supposed to capture gradient detail, it 'should not contain quantitative information. The proper level at which to describe variability of [alveolar] closure is actually the phonetic level.' (p.282) Scobbie (1995) makes a similar point in a different context when reviewing Zsiga's work on /s/ to /j/ assimilation (1995, discussed also in section 1.3.5) Zsiga found that unlike lexical palatalisation (e.g. which is obligatory and categorical at the lexical level as in *confess/confession*), post-lexical palatalisation (/s#j/) is gradient e.g. *...press your point...*[sʃj]. Scobbie *provisionally* argues, however, that a phonological assimilation rule can technically still be involved in the post-lexical palatalisation case. This yields a structure with a *contour segment* where a root node

(x) is allowed to dominate two opposite-valued binary features in a linear sequence on a single tier, as for the feature [anterior] here:

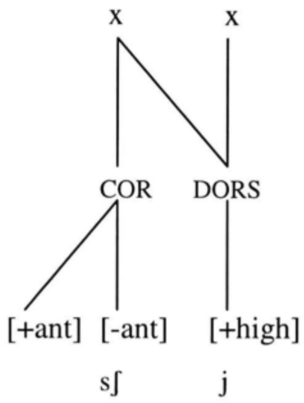


Figure 5.2 contour segment with a root node dominating two opposite-valued binary features (from Scobbie, 1995:310)

While in the case of lexical palatalisation [-ant] spreads from /j/ and in the process deletes [+ant] which is a feature of /s/, in the case of postlexical palatalisation, Scobbie argues, [+ant] and /j/ are not deleted: 'The first segment...has gradient anteriority and is symbolised as /sʃ/. The second, a dorso-coronal, can be symbolised as /sʃj/ or /jsʃ/...The target will, just as in the phonetic account, be subject to coarticulation.' (p.310) The important point being made here is that phenomena which appear to be temporally and spatially gradual, i.e. arising from coarticulation, is not necessarily predicted to be phonetic. Although new sub-phonemic categories are created, they 'have geometric properties which replicate aspects of the fine-grain temporal detail required by phonetics.' While describing how phonological formalism can technically express both categorical and gradient data, Scobbie believes this should be constrained. And here we are reminded of Hayes' objections to the modification of discrete phonological assimilation rules to capture the quantitative phonetic assimilation of Nolan's data (1992) which, he believes, should be handled at the phonetic level. Scobbie's solution in the case of palatalisation is to ban contour segments on the grounds that they are rarely used and that all they do anyway is 'to recapitulate non-contrastive, predictable and fine-grained temporal detail in surface structure.' (p.311) Similarly, Liberman and Pierrehumbert (1984) have advanced the view that all assimilatory processes should be treated as 'facts about the phonetic realisation of phonological representations rather than modifications of phonological representations themselves.' (p.166)

Hayes proposes that there should be two distinct processes to explain gradual assimilation: a phonological spreading rule, Place Assimilation and a phonetic rule, Alveolar Weakening. Hayes stipulates that the Place Assimilation rule produces not an entirely delinked place node for the coronal but the 'corono-dorsal complex segment' (contour segment). One of the advantages of Hayes' output is that it can express the idea that the alveolar and velar gestures may be prepared together, which this study provides some EMA evidence of. Marchal's proposal (1988) that stop consonant sequences are produced as a cohesive unit and not linearly gives some support to the underlying complex segment. This is certainly the case for consonant clusters produced by two independent articulators. Stetson (1951) showed that the movements required for clusters such as /pla/ or /fla/ are prepared together.

Once Hayes' underlying complex segment is in place the alveolar weakening rule applies. A problem lies in the fact that the output of the Place Assimilation rule, a complex segment, specifies that at some level in the generation and execution of the sequence an alveolar is always present. On the face of it, this is fine for analysing residual forms but is inappropriate for analysing the complete assimilations produced by 6 of the 10 speakers in the present study. In these cases there needs to be a model that predicts an absence of a coronal gesture. Hayes' formalism must be rejected since there ultimately seems no good reason to invoke a phonological rule *as well* as a phonetic rule where gradient assimilation is concerned and even if the place assimilation rule is posited independent of the phonetic rule, then the prediction (a double articulation) is wrong.

It is tempting to pursue the idea that *all* speakers (at least, in this study) have some commonality of phonetic/phonological information relating to assimilation but that at some other level, processing diverges for some speakers, which results in a different type of assimilation compared to others. It is likely that the motivation to establish such common ground comes from the tendency of successive theories of speech production to propose unified accounts of phenomena like assimilation. For instance, the traditional conception of alveolar to velar place assimilation itself is that it is a phonological phenomenon which later gave way to the idea that it is not phonological after all but phonetic. It has been shown in the last few paragraphs that assembling and combining rules of various sorts in order to analyse gradient and categorical data (since it must be accepted that this type of place assimilation is not always either one or the other) as underlyingly phonological, does not lead to anything like satisfying accounts.

Assimilation is not a universal phenomenon arising from inherent physical properties of the articulators since not all languages permit assimilation. And this means that the phonetic planning stage of speech production must be a more detailed mechanism than one which just assigns physical targets to more abstract specifications, for some speakers. The obvious question to ask at this juncture is: How does this discrepancy in processing path for assimilation come about? Perhaps a completely different approach can be suggested here which takes into account the wider sphere of speakers' language behaviour incorporating production, perception and early language experience. It may be the case that the mental storage of two discrete allophones of a phoneme (in this case /n/), and thus the production of categorical assimilatory forms as opposed to gradient reductions, is generated through the particular 'level of attention' a speaker pays to surface phonetic detail. In this way speakers with a perceptual balance of focus towards the phonetic level as apart from the phonological level may be more aware of perceived allophones (such as [n] and [ŋ] as allophones of /n/ when it occurs before /k/). 'More aware' would mean able to hear the difference between these optional allophones and to relate this to the speaker's production strategy, that is, there is a closer relationship between their own and others' speech production. This perceptual weighting (which may relate to awareness of tactile feedback) allows the speaker to abstract phonemic categories into separate stored allophones. According to speech style/rate, phonetic context and other constraints, these allophones could be optionally produced. For other speakers, on the other hand, the balance of focus may be different. These speakers may operate on a 'higher' level where there is less awareness of surface phonetics. This means that they may have no awareness of phonetic categories beyond /n/ and thus their production of variants on it will arise purely from the inertial effects on the tongue tip in certain speech styles/rates. Because of this assimilation will be gradient. This speculative approach to assimilatory behaviour has come about through the observation of how undergraduates learn phonetics and more specifically the ease with which surface contrast such as the context-bound plural allomorphs /s/, /z/, /ɪz/ can be perceived from which phonological generalisations can be made. The other intuition which this approach satisfies is the considerable variation in early language experience that infants receive. Quite apart from the difference in acquisition of production and perception skills that a deaf child experiences compared to their normally developing peers or that a normally developing child with hearing impaired parents experiences or a blind child experiences, within the normally developing population there must be ample scope for complex maturational interactions between the formation of phonological information and development of speech production awareness.

This approach may not stand up to generalisations covering other aspects of speech production or even the production of other connected speech processes and it does skate over some complex issues about production and perception, but it does rightly move away from too narrow a focus on the contents of psycholinguistic components which can only be the endpoint of wider influences on speech processes.

5.3 QUESTION 3: How do the articulatory details of assimilation match the existing models?

This section reviews three theoretical frameworks which have relevance to the issue of explaining connected speech processes. Each framework will be assessed in terms of its ability to accommodate the findings of the present study. Frameworks reviewed are Articulatory Phonology, Gesture Economy and the 'Windows' model of coarticulation.

5.3.1 Articulatory Phonology

The advantage of this model is that it naturally accommodates, and indeed predicts low-level, non-discrete phonetic detail in the case of place assimilation. It is proposed that it is primarily the sliding in time and hence overlapping of the alveolar and velar gestures on separate oral tiers which is the primary basis for fast/casual speech forms. Indeed, the time component of the Articulatory Phonology framework is absent from the other approaches evaluated here. Fast speech forms of this sequence produced by the gradual assimilators are unequivocally due to phonetic more-or-less adjustment in magnitude of the alveolar gesture but there were no cases of articulatory 'hiding' of the alveolar by the velar gesture said to be the result of maximally overlapping gestures and which, Browman and Goldstein assert, gives rise to the percept of assimilation. X-ray microbeam data of the utterance *perfect memory* presented by Browman and Goldstein (1990) shows how the perceived deletion of /t/ is not borne out by the articulatory facts - the /t/ gesture is fully present but overlapped by the /m/ (this data is discussed in section 1.2.2 of Chapter One). Browman and Goldstein did, however, go on to admit within the Articulatory Phonology framework (hereafter 'AP') the process of reduction in gestural magnitude. While it is easy to see how the sliding in time and hence overlapping of gestures on the gestural score is brought about by the compression of time it is difficult to see what mechanism governs reduction in magnitude and therefore why some segments should undergo reduction more than others. Browman and Goldstein argue that timing and amplitude interactions between gestures arise naturally out of the mathematical properties of the speech production apparatus.

Fowler and Saltzman (1993) have proposed the concept of *blending strength* to bring together the notions of coarticulatory resistance and coarticulatory 'aggression'. In this framework, all the elements of a gesture can be characterised by their own degree of blending strength which is derived from the demands individual elements place on the vocal mechanism. In a potential overlap situation a gesture with higher blending strength will suppress a gesture with a lower blending strength but when overlap occurs between gestures with equal blending strength status, the outcome is predicted to be the result of an averaging-out of the two effects. This is referred to as the task dynamic component of the model. Saltzman and Munhall (1989) state that in the task dynamic component, the vowels that surround this particular consonant sequence, an alveolar to velar sequence, share the same tract variable representation (tongue body) as the velar gesture. Thus the tongue body parameter is dominant in this VC_1C_2V sequence at the expense of the coronal gesture.

The fast speech /n#k/ data for the binary-option assimilators present the biggest problem of all for AP. AP claims to dispense with phonological rules altogether but it is difficult to see how a single pre-blended input to the gestural score can yield such optional and categorically different forms.

5.3.2 Gesture Economy

Gesture economy or ease of articulation seems to be invoked almost axiomatically as an explanation for place assimilation. While it provides no insights into the usual intractable questions surrounding assimilation such as why do alveolars and not velars assimilate, why does place assimilation occur from left to right but not from right to left and why do only some coronals assimilate and not all, gesture economy makes sense in view of the fact that the incidence of articulatory simplification (in the case of alveolar place assimilation, the place features for an alveolar stop changing into the place features for the following stop) increases as speech rate increases. This corresponds to Lindblom's duration dependent undershoot model (1963), but a possibility that might be considered is that complete assimilation is actually the default setting for the production of alveolar to velar sequences in normal speech. Physiological facts have been invoked to explain why this more economical predisposition might be in place. Saltzman and Munhall (1989) state that in their task dynamic framework of speech production, the vocalic gesture that contextualises this particular consonant sequence, an alveolar to velar sequence, share the same tract variable representation (tongue body) as the velar gesture and share the same extrinsic muscles. The tongue body parameter is then dominant in this sequence in relation to the coronal gesture. Viewed this way, the high number of complete assimilations produced in fast speech are accounted for while laboratory speech (less 'casual/conversational' and possibly self-conscious speech compared to unobserved normal speech) has introduced many non-assimilations in the speech of the 'non-assimilators' and also in the speech of 'binary categorical assimilators'. If, physiologically, alveolars are naturally predisposed to complete assimilation then why is it that they do not assimilate when they occur in a position such as /Vk#tV/? Clearly there is a perceptual constraint on assimilation which stipulates that in syllable-initial position an alveolar closure target, and thus release, must be preserved. Beginnings of words have priority over ends of words in English.

But how do residual alveolars relate to the concept of gesture economy? Although there is a reduction in the actual distance the tongue tip/blade has travelled in these cases, a motor command is none the less issued and in this sense there is no energy saving made nor is there a saving made at the planning level. If anything it is an expensive strategy because the articulators and supporting musculature have initiated an alveolar closure with no useful acoustic outcome. There is some evidence, however, that undershoot 'residues' of this sort may have some perceptual use in the recovery of alveolars (Nolan, 1992).

The problem with the concept of gesture economy, as Ohala (1990) argues, is that there is no clear definition of what constitutes articulatory 'ease' or 'simplification'. By coincidence, he gives the example of nasal place assimilation the result of which, he claims, is not necessarily simpler than the 'original' task since the velum is having to raise in the middle of, for instance, a bilabial gesture which spans two segments, rather than combine this raising gesture with a change in place of articulation (although the timing data from the present study shows that the offset of nasal voicing and the onset of velar closure is rarely simultaneous). As Ohala says, this may, in fact, be more taxing on the neurological commands. A further point against gesture economy is that in stop sequences it may be more economical to produce and maintain C_1 than C_2 rather than the other way around.

5.3.3 Windows model of coarticulation

In Keating's model (1990) there are a number of aspects which have particular relevance to the question of variability arising from place assimilation (see section 1.2.3). Unlike Articulatory Phonology, this framework considers that 'the phonological feature values that are the basis for window selection need not be the same as the underlying values: phonological rules to change or spread feature values still apply before the phonetics. Thus in terms of segments, windows are selected for extrinsic allophones rather than phonemes.'(1990: 456) At the same time, underspecification can be permitted to persist into the phonetic representation. The advantage of this heterogeneous framework is that discrete and continuous assimilatory effects are accommodated. The persistence of the unspecified coronal feature into the phonetic representation can account for the continuous assimilations produced by the gradual assimilators and a pre-Windows feature changing rule can account for the 'narrow window' categorical assimilations observed. This approach does not, however, explain why the unspecified item survives into the phonetics for some speakers and not others.

The other point in favour of Windows is that graded coarticulatory variation is argued to be the result of principled interaction between phonologically determined values and not universal and automatic characteristics of the vocal apparatus. Since it is clear that place assimilation is far from universal (see section 1.4 for studies of this), it makes sense to invoke a principled distinction between biomechanical variation and 'grammatical' variation, an approach lacking in other models.

5.4 CONCLUSIONS

The primary goal of this thesis has been to examine, on the basis of articulatory data from 10 speakers, the role of speech rate in the assimilation of alveolars which precede velars and to gain a clearer understanding of assimilatory variation. Regarding the effect of speech rate on assimilation, results showed that while assimilation is more common in fast speech than in slow/careful speech overall, speaker-specific preferences for non-assimilation can override this relationship. The correlation between rate and assimilation is further qualified by the finding that for some speakers, assimilation occurs in careful speech, indicating that assimilation is not determined by rate alone. In terms of the effect of rate on intergestural timing of events for non-assimilations, no dominant pattern emerges although there is a tendency towards a higher incidence of simultaneous alveolar release and velar closure in fast speech compared to careful speech. This may be to avoid the risk of the alveolar release occurring *after* velar release which would result in a hidden and inaudible velar gesture. Both across and within speakers there is considerable variability. Timing analyses also showed that assimilation is not the automatic consequence of rate-induced changes in intergestural timing of /n#k/. Assimilation is not triggered when overlap of alveolar and velar stop closure phases reaches a certain degree, nor is it triggered when the interval between onset of alveolar stop and velar stop becomes very short. It can be concluded from this that place assimilation has a separate origin from affects that may take place when physical movements compete for space and time.

Regarding the issue of variability in assimilatory forms, inter and intra-speaker differences are considerable. In fast speech, some speakers assimilate more than others. However, the most significant result was that some speakers assimilated gradually while others assimilated categorically. A very surprising result was that there were only two residual alveolar articulations (according to the classificatory criteria adopted in this study) in the entire database. This strongly conflicts with not only the general current approach to assimilation which assumes it is gradual but with the results of previous studies, for instance Kühnert's, where she states that all speaker produce partial assimilations at least some of the time. The interpretation made from the evidence of two distinct forms of assimilation is that some speakers utilise a cognitive assimilation rule while for others assimilation is 'computed on-line' during speech production itself. For the latter speakers, a phonetic weakening rule permits inertial effects on the articulators to the extent that a gesture can be non-existent. A speaker-specific basis for assimilation is proposed, which arises from variation in psycholinguistic processing of assimilation.

Phonological modelling of assimilation is ultimately unsatisfying in capturing both categorical and gradient assimilations. The notion of the phonological 'complex segment' where the coronal and dorsal features are bonded but still retain their original features, specifies that at some level in speech, planning a coronal gesture is present. This is clearly not the case for the completely assimilated /n#k/ tokens found in this study. Of the models of speech production reviewed, the Windows model provides at least the potential to account for both gradient and categorical assimilation. Phonological rules to alter underlying values can apply prior to window selection and yet underspecified values need not always be 'filled in' by rules. Articulatory Phonology copes well with the gradient assimilatory effects found in some of the data presented in this thesis, but cases of complete disappearance of the alveolar gesture are treated in a rather unsatisfying ad-hoc manner that edges away from the unifying principle of gestural blending. A more serious problem is that all gestures are treated as formally identical entities, leaving no opportunity to express coronals as phonologically special.

5.5 FUTURE DIRECTIONS

This study has highlighted that there is still much to be done to arrive at satisfactory modelling of alveolar to velar place assimilation data. In particular, attention needs to be paid to where in the generation and execution of an alveolar to velar sequence the gradual and categorical assimilators diverge and whether this has anything to do with how phonological processes interface with phonetic processes.

Apart from the need to work on better models, more data is needed. A more extensive database, formed using the same combined EPG/EMA methodology, could include productions of potential assimilation sites involving different motor subsystems with the hypothesis that assimilation has the potential to be gradient depending on the level of motoric 'independence' systems involved have from each other. For instance, in the utterance *...bad marks...* the tongue tip/blade and the lips are articulators from distinct systems and so the completely assimilated form [bab mɑrks] in fast speech might be dominant. Any residual alveolar movement for /d/ would be detected by the EMA trace although, again, extent of tongue tip raising would have to be compared with that which occurs for control bilabial to bilabial sequences. Alternatively, sequences involving different areas of the same articulator as in *...nice shower...* may be predicted to be more often gradient than categorical since this sequence is produced by the same articulator. Of course we already know that assimilation is not so straightforwardly motivated by universal biomechanical factors because assimilation is not part of the grammar for some languages. But the overall aim of this research would be to build up a picture of how biomechanical factors, speaker preference and speech rate contribute to assimilation. Subject numbers could be expanded to 40 although the number of repetitions produced by each speaker would not necessarily need to be increased. 10 repetitions has proved to be sufficient to make firm judgements although any less than this would not be desirable.

Another aspect of assimilation that deserves experimental investigation is the idea that place assimilation across words is more likely if those words have a high frequency (Kaisse, 1985; Dressler and Wodak, 1982). For instance, in an utterance such as *...in my...* or *...I'd go...* involving very high frequency function words, such as a preposition and a modal auxiliary, assimilation may not only be extremely common, even in slow speech, but may also be more likely to be categorical. Conversely, it may also be the case that place assimilation is gradient in nonsense words since speakers will not have the ability to scan ahead and plan conventionalised articulatory simplifications as they would in utterances that are very familiar to them.

Any further work in this area should include a re-assessment of EPG and possibly acoustic data from previous studies of place assimilation to look again for the type of speaker-specific differences in assimilation strategy found in the present study. Only time constraints prevented the inclusion of such a re-evaluation in this thesis. A closer look at the distribution of partial assimilations is needed to see if these are only produced by particular speakers. Holst and Nolan found in their acoustic study (1995) of /s/ to /ʃ/ assimilation that it could be categorical *or* gradient but it would be interesting to look at their data to discover if all subjects produce a mixture of these acoustically defined types or whether subjects produce only one or other of these types.

Finally, another experiment would be useful to determine if the status of an alveolar stop as nasal or non-nasal has any effect on the type of assimilations speakers produce. Hardcastle (1994) has suggested that nasal alveolar stops are more susceptible to assimilation than non-nasal alveolars. The place of articulation for a nasal stop does not give much information due to the fact that nasal cues have been found to be weaker than non-nasal place cues and that nasal stops are easily confusable with each other (House, 1957; Malécot, 1956). For these reasons it may be the case that complete assimilation is predominant.

REFERENCES

Ali, L., Gallagher, T., Goldstein, J. and Daniloff, R. (1971) Perception of coarticulated nasality. *Journal of the Acoustical Society of America* 49, 538-40

Archangeli, D. (1988) Aspects of underspecification theory. *Phonology Yearbook*, 5, 183-207

Barry, M. (1985) A palatographic study of connected speech processes. *Cambridge papers in Phonetics and Experimental Linguistics*, 4, 1-16

Barry, M. (1988) Assimilation in English and Russian. Paper presented at the colloquium of the British Association of Academic Phoneticians, Trinity College, Dublin, March 1988

Browman, C. and Goldstein, L. (1989) Articulatory gestures as phonological units *Phonology,* 6, 201-251

Browman, C. and Goldstein, L. (1990) Tiers in articulatory phonology with some implications for casual speech. In Kingston, J. and Beckman, M. (eds.) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge: CUP, 341-376

Browman, C. and Goldstein, L. (1992) Articulatory Phonology: an overview. *Phonetica 94* 155-180

Brown, G. (1990) *Listening to Spoken English*, Longman; Harlow

Byrd, D. (1992) Perception of assimilation in consonant clusters: a gestural model. *Phonetica*, 49: 1-24

Byrd, D. and Tan C.C. (1996) Saying consonant clusters quickly. *Journal of Phonetics,* 24, 263-282

Catford, J.C. (1977) *Fundamental Problems in Phonetics* Indiana University Press, Bloomington

Chen, M. (1970) Vowel length variation as a function of the voicing of the consonantal environment *Phonetica* 22, 129-159

Chomsky, N. & Halle, M. (1968) *Sound Pattern of English* New York: Harper and Row

Clark, J. and Yallop, C. (1990) *An Introduction to Phonetics and Phonology* Oxford: Basil Blackwell

Cohn, A.C. (1990) *Phonetic and Phonological Rules of Nasalisation* UCLA Working Papers in Phonetics, Number 76, Department of Linguistics, UCLA, Los Angeles.

Costa, P.J. and Mattingly, I.G. (1981) Production and perception of phonetic contrast during phonetic change *Status Report on Speech Research SR-67/68* Haskins Laboratories 191-196 New Haven, Conn.

Daniloff, R. and Moll, K. (1968) Coarticulation of lip rounding *Journal of Speech and Hearing Research* 11, 707-721

Daniloff, R. & Hammarberg, R. (1973) On defining coarticulation *Journal of Phonetics* 1, 239-248

Dinnsen, D.A. (1985) A re-examination of phonological neutralization *Journal of Linguistics* 21: 265-79

Dressler, W. & Wodak, R. (1982) Sociophonological methods in the study of sociolinguistic variation in Vienesse German *Language in Society* 11, 339-370

Farnetani, E. (1990) V-C-V lingual coarticulation and its spatio-temporal domain In Hardcastle, W.J. and Marchal, A. (eds) *Speech Production and Speech Modelling* Netherlands: Kluwer Academic 93-110

Farnetani, E. and Busà, M.G. (1994) Italian clusters in continuous speech *Proceedings of the 1994 International Conference on Spoken Language Processing*, 1, 359-62, Yokohama, 18-22 September

Fourakis, M. and Port, R. (1986) Stop epenthesis in English. *Journal of Phonetics* 14: 197-221

Fowler, C. (1983) Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 1, 113-133

Fowler, C. and Saltzman, E. (1993) Coordination and coarticulation in speech production In Marchal, A. and Farnetani, E. (eds) *Language and Speech* (special issue on coarticulation) vol. 36 parts 2 &3 April-September 171-197

Gibbon, F., Hardcastle, W.J. and Nicolaidis, K. (1993) Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation *Language and Speech* 36: 261-277

Gimson, A.C. (1960) The instability of English alveolar articulations *Le maître Phonetique,* 113

Gimson, A.C. (1989) *An Introduction to the Pronunciation of English* London: Edward Arnold (4th edition)

Hardcastle, W.J. (1972) The use of electropalatography in phonetic research. *Phonetica*, 25, 197-215

Hardcastle W.J. (1985) Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, 4, 247-263

Hardcastle, W.J. (1994) Assimilation of alveolar stops and nasals in connected speech. In J. Winsdor Lewis (ed.) *Studies in General and English Phonetics in Honour of Professor J.D. O'Connor* Routledge, 49-67

Hardcastle, W.J. and Roach, P. (1977) An instrumental investigation of coarticulation in stop consonant sequences *Work in Progress* 1: 27-44 Phonetics Laboratory, University of Reading

Hardcastle,W.J. and Morgan, R.A. (1982) An Instrumental Investigation of Articulation Disorders in Children *British Journal of Disorders of Communication* 17, 1, Apr, 47-65.

Hardcastle, W.J. and Barry, W. (1989) Articulatory and perceptual factors in /l/ vocalisation in English *Journal of the International Phonetic Association*, 15, 3-17

Hardcastle, W.J., Gibbon, F. and Nicolaidis, K. (1991) EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics* 19, 251-266

Hardcastle, W.J., Gibbon, F. and Jones, W. (1991) Visual display of tongue-palate contact: electropalatography in the assessment and remediation of speech disorders *British Journal of Disorders of Communication* 26: 41-74

Hawkins, P. (1984) *Introducing Phonology.* London: Hutchinson

Hayes, B. (1992) Comments on chapter 10. In G. Docherty and D.R. Ladd (eds.), *Papers in Laboratory Phonology II Gesture, Segment, Prosody.* Cambridge: CUP 280-286

Holst, T. and Nolan, F. (1995) The influence of syntactic structure on [s] to [ʃ] assimilation. In Connell, B. and Arvaniti, A. *Papers in Laboratory Phonology IV* Cambridge: Cambridge University Press 315-333

Hoole, P. (1993) Methodological considerations in the use of electromagnetic articulography in phonetic research. *Forschungsberichte des Institut für Phonetic und Sprachliche Kommunikation der Univerität München* 31 3-175

Hoole, P. and Nguyen, N. (1992) Reduction and parameterization of electropalatographic and electromagnetic data for the investigation of articulatory-acoustic correlations *Forschungsberichte des Institut für Phonetic und Sprachliche Kommunikation der Univerität München* (FIPKM) 30, 45-72

House, A.S. (1957) Analog studies of nasal consonants *Journal of Speech and Hearing Disorders* 22: 190-204

Householder, F.W. (1956) Unreleased /ptk/ in American English. In M. Halle (ed.) *For Roman Jakobson* The Hague: Mouton, 235-244

Hura, S., Lindblom, B. and Diehl, R. (1992) On the role of perception in shaping phonological assimilation rules *Language and Speech* 35 (1, 2) 59-72

Jones, D. (1932) *Outline of English Phonetics* Cambridge University Press (3rd edition)

Jun, J. (1996) Place assimilation is not the result of gestural overlap: evidence form Korean and English *Phonology 13:* 377-407

Kaisse, E.M. (1985) *Connected Speech: the interaction of syntax and phonology* Academic Press, Orlando, U.S.A.

Keating, P. (1988) Underspecification in phonetics *Phonology* 5: 275-92

Keating, P. (1990) The window model of coarticulation: articulatory evidence. In Kingston, J. and Beckman, M. (eds.) *Papers in Laboratory Phonology I: between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press, 451-471

Keating, P. (1996) The phonology-phonetics interface. *UCLA Working Papers* 92: 45-60

Kelso, J.A.S., Saltzman, E.L., Tuller, B. (1986) The Dynamical Perspective on Speech Production: Data and Theory *Journal of Phonetics*, 14, 29-59

Kent, R. and Minifie, F.D. (1977) Coarticulation in recent speech production models *Journal of Phonetics,* vol. 5, 115-133

Kerswill, P. (1985) A sociophonetic study of connected speech processes in Cambridge English: An outline and some results. In Nolan, F.J. (ed.) *Cambridge papers in Phonetics and Experimental Linguistics* 4, 1-39

Kerswill, P. (1987) Levels of linguistic variation in Durham *Journal of Linguistics,* 23, 25-49

Kohler, K. (1976) The instability of word-final alveolar plosives in German: an electropalatographic investigation. *Phonetica* 33, 1-30

Kohler, K. (1990) Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In Hardcastle , W. and Marchal, A. (eds.) *Speech Production and Speech Modelling.* Dordrecht: Kluwer, 69-92

Kohler, K. (1992) Gestural reorganisation in connected speech: a functional viewpoint on 'Articulatory Phonology' *Phonetica* (commentary) 49: 205-211

Kühnert, B. (1993) Some kinematic aspects of alveolar to velar assimilations. *Forschungsberichte des Institut für Phonetic und Sprachliche Kommunikation der Univerität München* (FIPKM) 31, 263-72

Kühnert, B and Nolan, F. (1999) The origin of coarticulation In Hardcastle, W.J. and Hewlett, N. (eds.) *Coarticulation: theory, data and techniques* Cambridge University Press: Cambridge

Ladefoged, P (1990) On dividing phonetics and phonology: comments on papers by Clements and Browman and Goldstein. In Kingston & Beckman, M. (eds.) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge: Cambridge University Press 398-405

Ladefoged, P. (1993) *A Course in Phonetics* (3rd edition) Harcourt, Brace, Jovanovich: New York

Laver, J. (1994) *Principles of Phonetics* Cambridge: Cambridge University Press

Levelt (1989) *Speaking: from intention to articulation.* Cambridge, MA: MIT Press

Liberman, M and Pierrehumbert, J (1984) Intonational invariance under changes in pitch range and length In *Language and sound structure* ed. Mark Aronoff and Richard T. Oehrle Cambridge, Mass.: MIT Press

Lindblom, B. (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-81

Lindblom, B. (1983) Economy of speech gestures In MacNeilage, P.F. *The Production of Speech* New York, NY: Springer-Verlag, pp217-45

Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H & H theory. In W.J. Hardcastle and A. Marchal (eds.), *Speech production and Speech Modelling* 403-440 Dordrecht: Kluwer Academic Publishers

Lindblom, B., Pauli, S. and Sundberg, J. (1975) Modeling coarticulation in apical stops In G. Fant (ed.) *Speech Communication* , vol. 2, 87-94 Uppsala, Almqvist and Wiksell

Local, J. (1992) Modelling assimilation in nonsegmental, rule-free synthesis. In Docherty, G.J. and Ladd, D.R. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody.* Cambridge: Cambridge University Press 190-223

Lubker, J. F. & Gay, T. (1982) Anticipatory labial coarticulation: experimental, biological and linguistic variables *Journal of the Acoustical Society of America* 71: 437-448

Maeda, S. (1999) Labialisation during /k/ followed by a rounded vowel is not anticipation but the auditorily required articulation *Proceedings of the XIVth International Congress of Phonetic Sciences,* San Francisco, 1-7 August, 41

Malécot, A. (1958) The role of releases in the identification of released final stops. *Language* 34, 370-380

Marchal, A. (1988) Coproduction: evidence from EPG data. *Speech Communication* 7: 287-295

Nguyen, N. (1996) A Matlab package for the analysis of EMA/EPG data *QMC Working Papers in Speech and Language Sciences* (ed. J.M. Scobbie) 1: 5-12

Nolan, F. (1985) Idiosyncrasy in coarticulatory strategies *Cambridge Papers in Phonetics and Experimental Linguistics* 4: 1-9

Nolan, F. (1986) The implications of partial assimilation and incomplete neutralisation. *Cambridge Papers in Phonetics and Experimental Linguistics* Volume 5 1-11

Nolan, F., Holst, T. and Kühnert, B. (1996) Modelling [s] to [ʃ] accommodation in English *Journal of Phonetics* 24 (1) 113-137

Nord, L. (1986) Acoustic Studies of Vowel Reduction in Swedish *Speech Transmission Laboratory Quarterly Progress and Status Report*, Department of Speech Communication STL-KTH Stockholm, 19-36

O'Connor, J.D. (1973) *Phonetics* Harmondsworth: Penguin

Ohala, J. (1974) Experimental historical phonology In *Historical linguistics II Theory and description in phonology* (J.M. Anderson & C. Jones, eds.) 353-389 Amsterdam: North Holland Publishing Co.

Ohala, J. (1990) The phonetics and phonology of aspects of assimilation. In Kingston, J. and Beckman, M. (eds.) *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech.* Cambridge: Cambridge University Press, 258-275

Ohala, J. (1993) Coarticulation and phonology In Marchal, A. and Farnetani, E. (eds.) *Language and Speech* (special issue on coarticulation) vol. 36 parts 2 & 3, 155-171

Öhman, S. (1966) Coarticulation in CVC utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39, 151-168

Öhman, S. (1967) Numerical model of coarticulation *Journal of the Acoustical Society of America* 41, 310-320

Paradis, C. and Prunet, J-F (1991) *The special status of coronals* Dordrecht: Foris

Perkell, J.S., Cohen, M.H., Svirsky, M.A., Matthies, M.L., Garabieta, I. and Jackson, M.T.T. (1992) Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements *Journal of the Acoustical Society of America* 92, 3078-3096

Pierrehumbert, J. and Talkin, D. (1992) Lenition of /h/ and glottal stop. In D.R. Ladd and G.J. Docherty (eds.) *Papers in Laboratory Phonology II* 90-117 Cambridge University Press

Port, R. F. and O'Dell, M.L. (1985) Neutralisation of syllable-final voicing in German. *Journal of Phonetics* 13, 455-471

Recasens, D., Fontdevila, J., Pallarès, M.D., Solanas, A. (1993) An electropalatographic study of stop consonant clusters *Speech Communication* 12: 335-355

Roach, P. (1983) *English Phonetics and Phonology: a practical course* Cambridge: Cambridge University Press

Saltzman, E. and Munhall, K. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1, 333-382

Scobbie, J.M. (1995) What do we do when phonology is powerful enough to imitate phonetics? Comments on Zsiga In *Papers in Laboratory Phonology IV phonology and phonetic evidence* ed. B. Connell and A. Arvanti, Cambridge: Cambridge University Press 303-313

Scobbie, J.M., Coleman, J.S., Bird, S (1996) Key aspects of Declarative Phonology *Current Trends in Phonology: Models and Methods.* Durand, J. and Laks, B. (eds.) (2 volumes) European Studies Research Institute (ESRI), University of Salford: Manchester UK Volume 2: 685-709

Solé, M.J. and Ohala, J. (1991) Differentiating between phonetic and phonological processes: the case of nasalization *Proceedings of the 12th International Congress of Phonetic Sciences* 2, 110-113, Aix-en-Provence

Spencer, A. (1996) *Phonology: Theory and Description* Oxford: Blackwell

Sproat, R and Fujimura, O. (1993) Allophonic variation in English /l/ and its implications for phonetic implementation *Journal of Phonetics* 21: 291-311

Stetson, R. (1951) *Motor Phonetics: a study of speech movements in action* (2nd edition) Amsterdam: North Holland

Stevens, K.N., House, A.S., Paul, A.P. (1966) Acoustical description of syllabic nuclei: an interpretation in terms of a dynamic model of articulation *Journal of the Acoustical Society of America* 40, 123-132

Sussmanm H. M. and Westbury, J. (1981) The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation *Journal of Speech and Hearing Research* 24, 16-24

Walley, A. and Carrell, T. (1983) Onset spectra and formant transitions in adult's and child's perception of place of articulation in stop consonants *Journal of the Acoustical Society of America* 73, 1011-1022

Wells, J.C. (1982) *Accents of English* Vols. 1-3 Cambridge: CUP

Wickelgren, W. (1969) Context-sensitive coding, associative memory and serial order in (speech) behaviour *Psychological Review* 76: 1-15

Winitz, H., Scheib, M.E. & Reeds, J.A. (1972) Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech *Journal of the Acoustical Society of America* 51, 1309-1317

Wood, S.E. (1995). An electropalatographic analysis of stutterers' speech. *European Journal of Disorders of Communication,* Vol. 30, No. 2.

Wrench, A. A. (in press) A multi-channel/multi-speaker articulatory database for continuous speech recognition research. In Phonus, Research Report No. 4, Institute of Phonetics, University of Saarland

Wright, S. and Kerswill, P. (1989) Electropalatography in the study of connected speech processes. *Clinical Linguistics and Phonetics* 3, 49-57

Zsiga, E. (1995) An acoustic and electropalatographic study of lexical and postlexical palatalization in American English In *Papers in Laboratory Phonology IV phonology and phonetic evidence* ed. B. Connell and A. Arvanti, Cambridge: Cambridge University Press 282-302

Zsiga, E. (1997) Features, gestures and Igbo vowels: An approach to the phonology phonetics interface *Language* 73, no.2: 227-271

Conference Proceedings Publications

An EPG study of alveolar to velar coarticulation in fast and careful speech: some preliminary observations

Lucy Ellis & W. J. Hardcastle
Department of Speech and Language Sciences, QMC, Edinburgh, UK

1. Introduction

This paper is concerned with the articulatory details of a connected speech process - the assimilation of a word-final alveolar nasal to a following velar plosive, under conditions of varied speech rate.

A commonly observed phenomenon in a number of different languages is the so called 'instability' of alveolar sounds /t, d, n/. These sounds tend to assimilate to following velars and bilabials /k, g, p, b/ - that is, alveolars can take on the place of articulation characteristics of an adjacent segment and either change into or become more like it.

Traditionally, the phonological account of this coarticulatory phenomenon favours a simple binary description reflecting the perceived presence or absence of an alveolar: e.g. *tin can* /tɪn # kan/ → /tɪŋ # kan/. However, studies with electropalatography (EPG), a technique which records spatio-temporal patterns of tongue-palate contact, have established that this description does not capture the full range of patterns that actually occur. Hardcastle and Roach (1979) and Wright and Kerswill (1989) looked at articulation of alveolars in cluster sequences and found that a continuum of patterns across subjects occur ranging from full alveolar closure to 'complete' assimilation (no evidence of tongue-tip/blade contact or lateral contact beyond that which characterises an EPG description of a lexical velar). Varying degrees of 'residual' or 'partial' assimilations occupy the area in between. Hardcastle (1994) found that alveolar nasals are more susceptible to assimilation than plosives. The assimilatory behaviour of alveolar nasals remains a relatively under-researched area.

Nolan (1992) tested the perceptual response of listeners to this continuum of articulations in order to discover what the perceptual correlates of articulatory gradualness might be. He found that while the identification of complete alveolar closure with lexical alveolars is highly reliable, the identification of residual alveolars with lexical alveolars is ambiguous. Furthermore there is no conclusive evidence that listeners are able to recover an alveolar from a 'completely' assimilated alveolar-velar sequence. However, when naive listeners were asked to identify these when presented as a pair with lexical velars, they scored rather better than when presented with them unpaired. If listeners are to some extent able to match a residual phonetic trace of an alveolar stop to their mental lexical representation of that place (and therefore 'restore' it), then it would be reasonable to suggest that speakers too have a more detailed mental lexical representation than is otherwise assumed.

It has long been suggested that differences in lexical phonological form such as 'assimilatory' [ŋ] and lexical [ŋ], will always result in distinct articulatory or acoustic forms, at least for this type of optional assimilation (Nolan, 1992, although the author has since distanced himself from this hypothesis). Reliable evidence for this in EPG studies is not forthcoming although it is possible that studies using alternative instrumentation may establish the existence of a utilisable trace of the alveolar at some level. If not, the prospect of gestural deletion in this context provides a challenge to the theory of Articulatory Phonology where it is claimed that place assimilation is mainly the result of gestures overlapping and the perception of these as a single gesture. For instance, Browman and Goldstein (1990) consider a casual realisation of the phrase *hundred pounds*. They propose that if the /p/ is superimposed on it, the /d/ gesture is still maintained: 'The bilabial closure gesture may increase its overlap with the preceding alveolar gesture, rendering it effectively inaudible. The overlap of voicing onto the beginning of the bilabial closure yields the [bp] transcription' (p.361). Jun (1996) in his study on place assimilation in *pk* clusters in Korean and English argues that gestural overlap cannot be the sole factor behind place assimilation. He also argues that gestural *reduction* 'is speaker-controlled; it does not result directly from physical constraints on speech-production mechanisms' and therefore is the source of the variability in place assimilation.

Variation in speech rate/style is known to have an effect on assimilation. Other influencing variables include syntactic structure, stress and informational load. Rate has a more significant effect on coarticulation than syntax (Hardcastle, 1985) and a general if not unsurprising finding common to all research in this area is that coarticulation and connected speech processes tend to be applied at fast rather than slow speaking rates. The other tendency is for speaker-specific strategies (different responses to the demands on the articulators of increased rate) to emerge. This dimension of assimilation has only been explored with subject groups as small as two or three - the study reported here is a more systematic, larger scale effort. The robust correlation of rate with connected speech processes is not, however, straightforward. Brown (1990) has demonstrated in an auditory study that assimilation does occur in a 'careful' style of speech. Also, Barry (1985) and Kerswill (1985) using EPG found that while at a faster rate subjects tended to make less alveolar contact, when required to speak fast but 'carefully' this tendency could be overridden.

Durational variations can occur when an articulatory organ does not have sufficient time to complete a given target and so has to 'undershoot' it. This is the basis of Lindblom's 'duration dependent undershoot' model (1963) which was developed from acoustic evidence of vowel reduction. He proposed that articulatory and acoustic undershoot of vowels is a function of *reduction of movement* towards the vowel target due to physiological limitations. This assumes though, that segments in connected speech are reduced equally in duration at fast rates. Another articulatory outcome is an increase in movement velocity (Lindblom, 1990) although this can combine with undershoot (spatial reduction) to give rise to a further strategy (Gay, 1981).

Place assimilation behaviour is not a universal phenomenon. Lindblom (1983) views assimilation as a language specific grammatical rule which represents a categorical change unlike coarticulation which is a continuous motor process. Provisional confirmation of the view that alveolar-velar assimilation in Russian is nowhere near as extensive as it is in other languages, has been provided in an EPG study by Barry (1988). Most interestingly, Farnetani and Bùsa (1994) found that in Italian the alveolar-velar assimilation in /nk/ clusters is always categorical. On the basis of an auditory study of Durham English, Kerswill (1987) notes either an absence or near absence of place assimilation in contexts where it might reliably be predicted to occur in other varieties of English. Once the distribution of place of assimilation behaviour across many languages and accents comes to light we can fully conceive of it as something over which speakers have control. This could have ramifications for theories that are concerned with the precise level of phonetic detail that is specified in speakers' mental representations or phonetic plans.

2. Method

2.1 Stimuli

Speech material was devised so that consonantal combinations would capture potential sites of alveolar assimilation in addition to neutral velar control contexts. These experimental combinations were embedded in the sentences: *"I can't believe the ban cuts no ice "* /n#k/ and *"I've heard the bang comes as a big surprise"* /ŋ#k/. Further material was devised to capture two other consonantal combinations: /n#t/ *"I'm not surprised the ban touched a raw nerve"* and /ŋ#t/ *"I reckon the bang toughened her resolve"* the results from which are not reported here.

The vocalic environment was kept as consistent as possible. Bordering vowels were /a/ & /ʌ/ and the /a/ vowel was preceded by a bilabial stop to eliminate the possibility of any lingual coarticulatory effects on the target consonants. This latter control will be particularly advantageous in the light of a planned EMA experiment using similar test stimuli where the only source of coarticulation on tongue trajectories will be the tongue tip and dorsum themselves. Low vowels were used to flank the consonants because vocalic tongue-palate contact (characteristic of high front vowels for instance) would be minimal. Also it was predicted that a rather 'front' velar occlusion would be achieved after /a/, more easily observed on EPG printouts. Voiceless plosives were selected to follow the nasal in all experimental sentences so that the offset of voicing for the nasal could be directly measured in relation to the transition from nasal stop to oral stop. This is especially useful in those cases where phonetic differences are sought between 'assimilatory' /ŋ/ to /k/ and lexical /ŋ/ to /k/, *...ban cuts.../...bang comes...*, even though these are not strictly minimal pairs.

A further 4 'filler' sentences of no experimental interest were added to the original 4 to distract the speakers from the presence of near minimal pairs. 10 repetitions of each test item were produced.

2.2 Speakers

10 speakers with EPG palates were recorded. All but 2 of the subjects were female and overall the subject group represented a fairly wide range of regional accents. 4 of them spoke what might be called Standard Southern British, one was Australian, one was Northern Irish, one spoke a variety of North Eastern English and the remaining 3 spoke Scottish English from 2 different regions.

2.3 Data collection

The technique of electropalatography (EPG) was used to record the timing and the location of tongue contact with the hard palate during continuous speech. During the course of the recording each speaker wore an artificial palate embedded with 62 silver electrodes which are activated when contact occurs. The details of contact are then stored on computer. The palate can be roughly divided into three zones: alveolar region (rows 1 and 2), palatal region (rows 3-5) and velar region (6-8). Although palates are custom made for speakers these regions follow predetermined anatomical landmarks which target phonetically significant areas. The sampling rate of this system is 100 frames per second and the acoustic signal is 10kHz.

Before the experiment speakers wore their palates for an hour to adjust, although they in fact required little acclimatization. Most of the subjects were experienced wearers of EPG palates and had been involved in other experiments using the technique.

The experiment fell into two parts. The aim of the first part was to elicit careful speech and all subjects were instructed to read each sentence *slowly and clearly*. No particular instruction was given with respect to prosodic phrasing. The aim of the second part was to elicit fast/casual speech although genuinely casual speech is notoriously difficult to acquire under laboratory conditions. The material for this part was identical to the first but the 80 sentences were arranged in groups of 3 and filler sentences were purposely distributed to avoid 'clustering' of experimental sentences. Subjects were instructed to read out each group of sentences in a *rapid and casual style* one after the other avoiding obvious pauses in between each sentence. A time limit of 5 seconds was imposed on the delivery of each group of sentences and time was allowed for subjects to practice both the fast groups and the individual careful sentences. The time constraint automatically ruled out any attempt on the part of speakers to impose too complex an intonational structure on each sentence. All subjects perceived this time limit as a challenge so it was an effective measure against over-awareness of the test items. The sentences in the first half of the experiment and the groups in the second were individually cued during recording with a pause between each.

2.4 Data analysis

Electropalatographic measurements were taken from the EPG trace and acoustic measurements were taken from the waveform/spectrogram. For all experimental sequences annotation points were made from the onset of the vowel before the word boundary (/a/ in all cases) up to onset of the vowel after the word boundary (/ʌ/ in all cases).

5 annotation points were made for sequences where a single place of articulation was achieved i.e. /ŋ#k/ *...bang comes...* in fast and careful rate and /n#k/ *...ban cuts...* in the fast rate, where there was a complete absence of alveolar contact or it was insufficient to justify annotation. It was also predicted that, for some speakers, a careful articulation of this latter sequence would motivate an assimilation of this sort. 7 annotation points were made for sequences where there were 2 clearly observable adjacent places of articulation. That is, in the context /n#k/ in the careful and fast speech rate where the alveolar target was achieved. A database was created which stores the EPG frame corresponding to each annotation point and its time value.

Figure 1 below shows a waveform, spectrogram and EPG display of one canonical repetition of *...ban cuts...* in the careful speech condition. The numbered annotation points are indicated on the spectrogram and the timing of these in relation to tongue-palate contact during the sequence is marked on the EPG display below (in each frame of EPG data the top of the schematic plan is the alveolar ridge and the bottom row approximately corresponds to the junction between the soft and the hard palate). These annotation points are defined in Table 1 below.

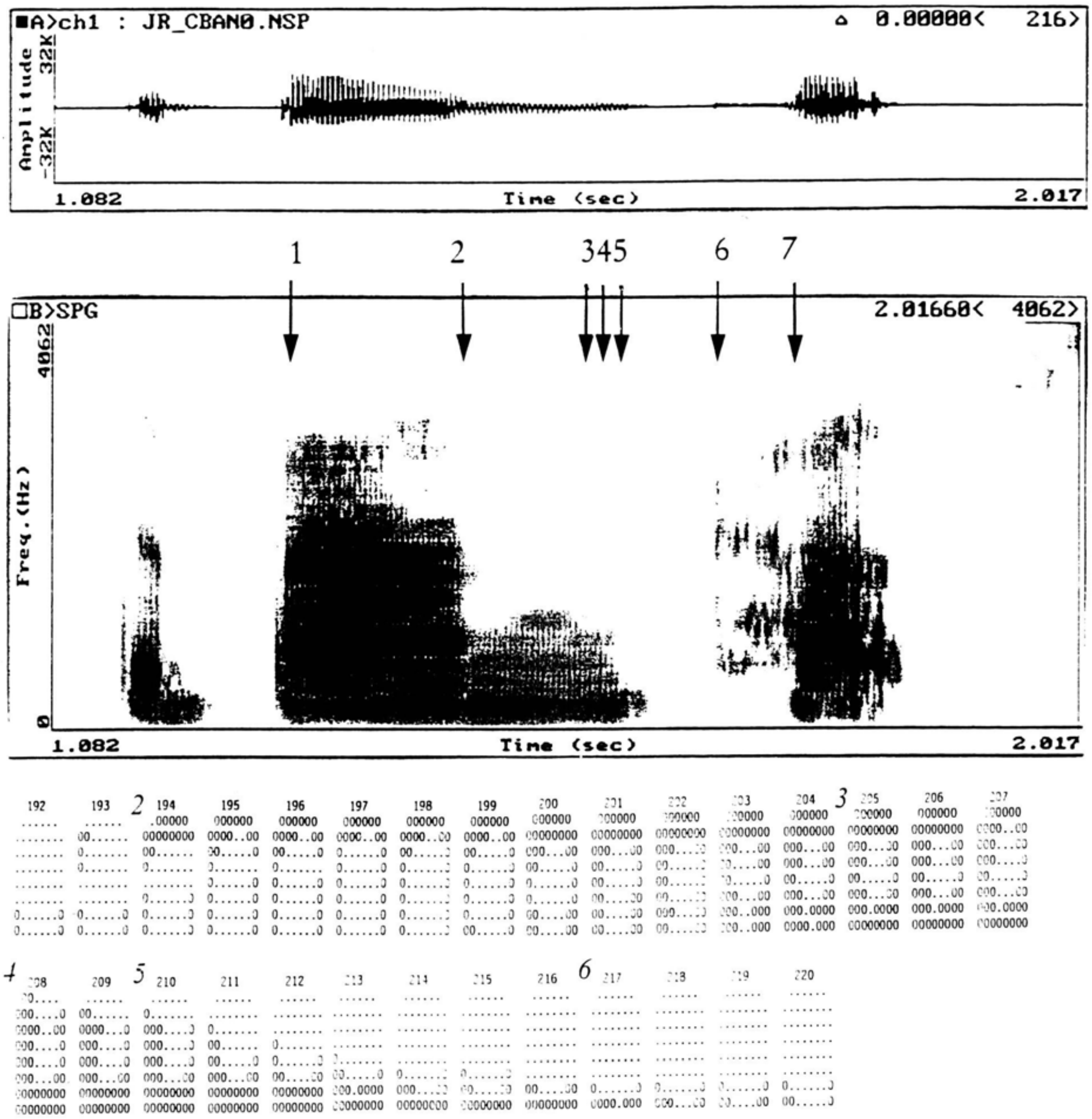


Figure 1: waveform, spectrogram and EPG display of...*ban cuts*... careful speech (subject JR). Numbered arrows represent acoustic and EPG-defined annotation points.

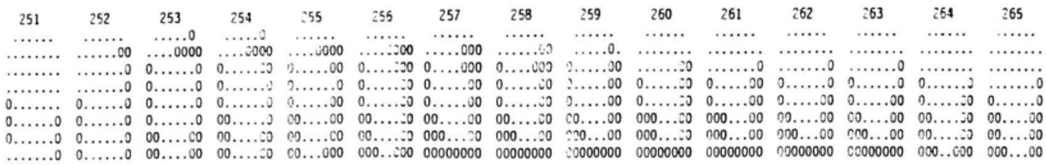
TABLE 1. Definition of annotation points: acoustic and EPG

no.	type	definition
1	acoustic	Onset of periodicity for the vowel /a/
2	EPG	Onset of mid sagittal contact in first 3 rows for alveolar /n/
3	EPG	Onset of complete or maximum constriction in row 8 for velar /k/
4	EPG	Earliest appearance of loss of contact for release of /n/
5	acoustic	End of nasal formant structure for /n/
6	EPG	Earliest appearance of loss of constriction for velar /k/
7	acoustic	Onset of periodicity for the vowel /ʌ/

N.B. The order of annotation points as they appear in the example in Fig.1 & Table 1 is, obviously, subject to variation. For example, formant structure for /n/ can end before the alveolar closure is released.

For analysis of the /n#k/ data a clear distinction had to be made between a 'canonical' alveolar or allophone of an alveolar and an assimilation of an alveolar. For a preliminary indication of the assimilatory trends, partial assimilations were subsumed into the general category of assimilation. To be classed as an alveolar, a pattern had to show mid sagittal contact in the first three rows of the EPG palate. Thus according to this definition, in Figure 2 (a) we see an allophone of /n/ while in (b) we see a partial alveolar articulation which here would be labelled an allophone of /ŋ/.

2(a)



2(b)

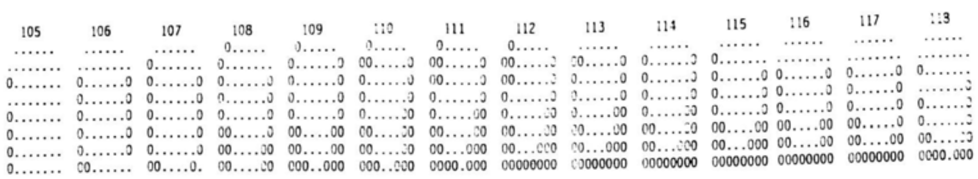
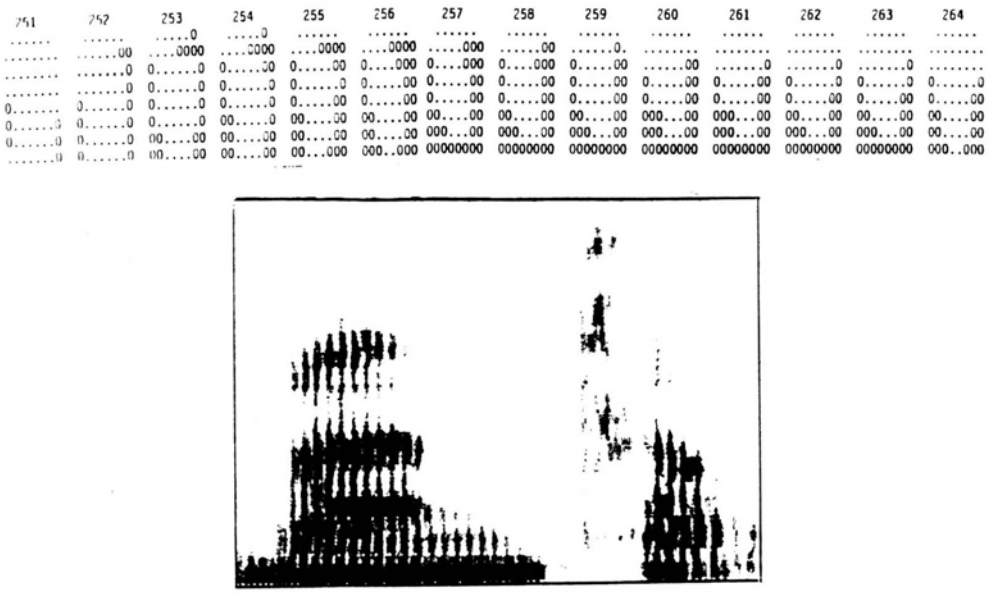


Figure 2 EPG patterns for...*ban cuts*... fast speech 2(a) annotation of an alveolar stop (JSC) 2(b) no annotation of an alveolar stop (WCM)

Although Fig.2 (b) is not classed as an alveolar articulation, the tongue has made the supporting lateral gesture for /n/ and as a result of tongue-tip raising, an acoustic correlate similar to one of a full alveolar might be expected. Figure 3 shows spectrograms for (a) a partial closure at the alveolar ridge and (b) a complete assimilation for *...ban cuts...* fast rate. F2 in (b) stays level while in (a) it drops slightly as the tongue tip makes contact. Notice also the lack of transition in (b) for the velar constriction. More thorough acoustic analysis of complete assimilations in this context and their comparison with lexical velars is planned.

3(a)



3(b)

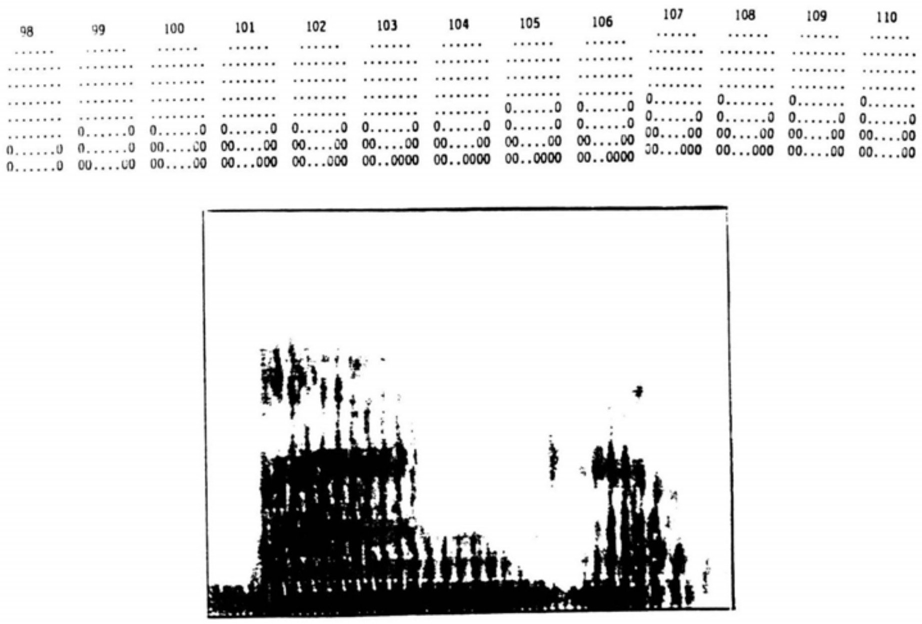


Figure 3 Spectrograms and EPG patterns for (a) incomplete alveolar closure (b) assimilated alveolar. Subject JSC

3. Preliminary Results and Discussion

The distribution of assimilations for /n#k/ in both fast and careful speech for all subjects are shown in Table 2.

TABLE 2

	careful speech	*fast speech*
assimilated	**3**	**56**
non-assimilated	**97**	**44**

The first thing to notice is that there are surprisingly few assimilated alveolars in the careful speech condition. These were produced by 2 subjects JF and SW who rather appropriately happened to be subjects who produced 100% assimilations of the same sequence in the fast rate. This might suggest that their systems are more predisposed to connected speech processes than others. The other thing to note from Table 2 is that although there are more assimilations in the fast rate as might be predicted, they by no means dominate the picture.

A breakdown of the results for fast speech /n#k/ for each subject is shown in Figure 4 below. There is considerable gross variation between speakers with regard to occurrence of assimilation. 2 subjects never assimilated in fast speech and 3 subjects always did (one subject assimilated for 9 out of 10 repetitions). The picture becomes more complex, however, if the patterns for the subjects in the middle of the graph are considered in more detail. Subjects FG, WCM, JSC, TH sometimes 'assimilated' and sometimes did not. But the graph does not tell the whole story, of course, since partial assimilations are not represented here. While the values for subjects JSC and TH appear the same on the graph they do in fact use entirely different reduction strategies as interpreted from the EPG contact patterns.

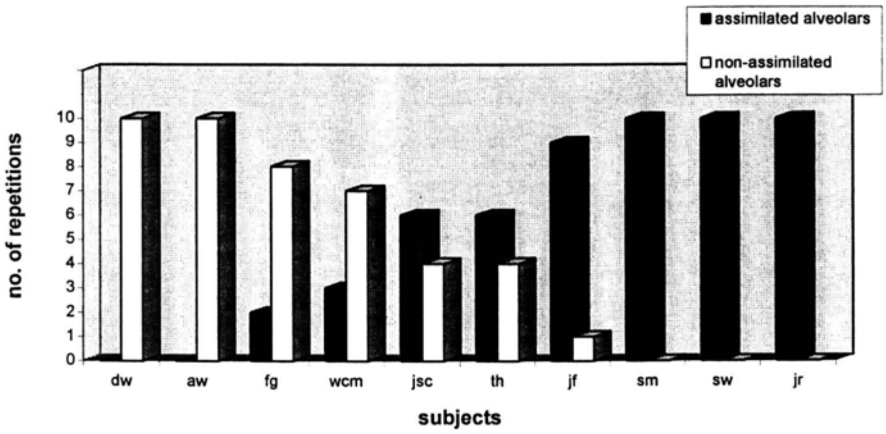


Figure 4 Distribution of assimilations for all 10 subjects ...*ban cuts*...fast speech

The EPG patterns for TH and FG suggest the adoption of a binary segmental strategy. That is, either full alveolar contact is achieved for target /n#k/ or there is segmental substitution with something that looks very similar to /ŋ/. The latter appears to be a restructuring option that precludes gradient partial alveolar assimilations. But partial assimilations *are* tolerated by JSC and WCM. Figure 2 shows the type of 'undershoot' articulations made by these subjects.

On the basis of this observation it would appear that not all speakers have the same 'allophone tolerance' for /n/ in this phonetic environment and in the articulatory domain at least. TH and FG behave here as speakers whose mental representation of /n/ specifies only a single alternative articulatory target. JSC and WCM are, however, speakers for whom permissible realisations are not categorical but include presumably unlimited graded intermediate articulations. One wonders how the contents of a mental lexical representation for these speakers might be expressed. It was thought possible that the discreteness of the two groups in this respect is a function of one of the groups speaking at a faster rate than the other. Target undershoot as a possible manifestation of inertial effects on the articulators is probably more closely associated with fast speech rate than segmental substitution, since the latter is known to occur even at careful rate. Measurements of duration of the sentences did not show JSC's and WCM's fast speech rate to be significantly faster than that of TH and FG (see Figure 5 below). Furthermore, the variability range for the fast speech repetitions for one group is not broader or narrower than for the other.

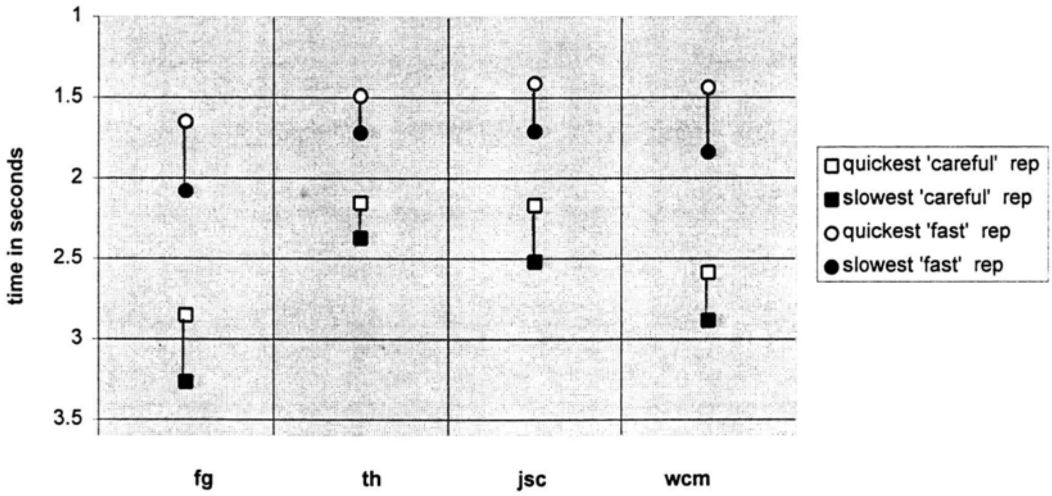


Figure 5. Duration ranges of careful and fast speech sentences for 4 subjects.
N.B. quickest utterance times at the top end of the graph.

Residual alveolars of the type that showed up in the data for JSC and WCM were almost completely absent in the data for the other 8 subjects. JSC and WCM were the only speakers whose alveolar-velar articulations were clearly gradient. They are, incidentally, both from West Scotland.

Figure 6 shows the EPG patterns for all 10 repetitions of *...ban* cuts...fast rate, for subject JSC. Each line shows a single repetition and captures the realisation of /n#k/. They are ordered to show a gradation from full alveolar contact to complete assimilation via consonant shortening of /n/ and target undershoot.

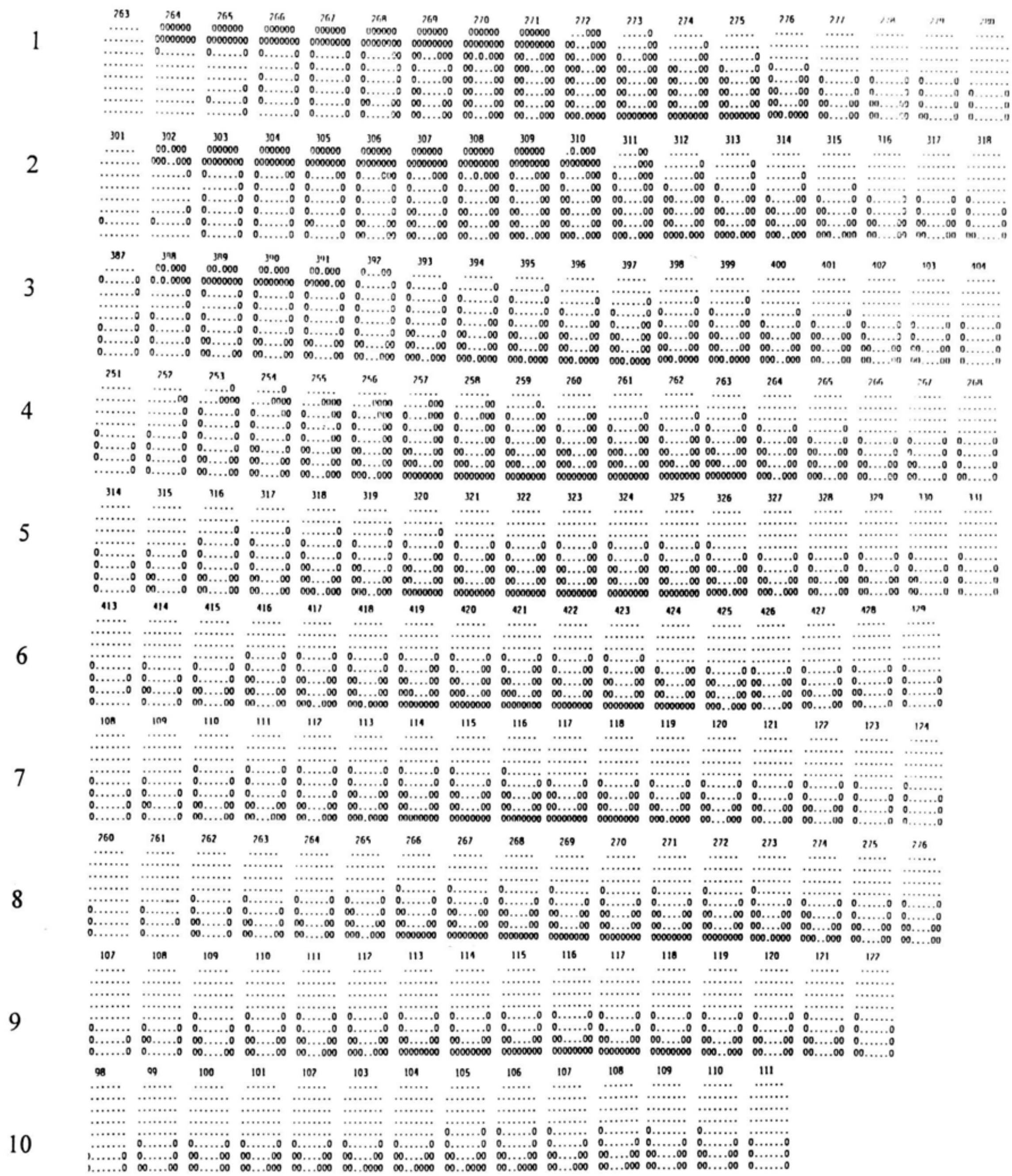


Figure 6 EPG patterns for all 10 realisations of /n#k/ sequence in *...ban cuts...*fast speech. (JSC)

The EPG contact 'totals' profiles for the patterns shown in Figure 6 are plotted in Figure 7 (a) below. Contact totals for all 10 repetitions from the beginning of the vowel /a/ are superimposed on a single graph. For each repetition there are two curves representing the amount of electrodes contacted on the palate frame by frame in the alveolar region (front 3 rows) and the velar region (back 3 rows). 7 (b) shows 10 realisations of the sequence...*ban cuts*...for fast speech by subject AW. 7(c) shows 10 realisations of the same sequence by AW but for careful speech.

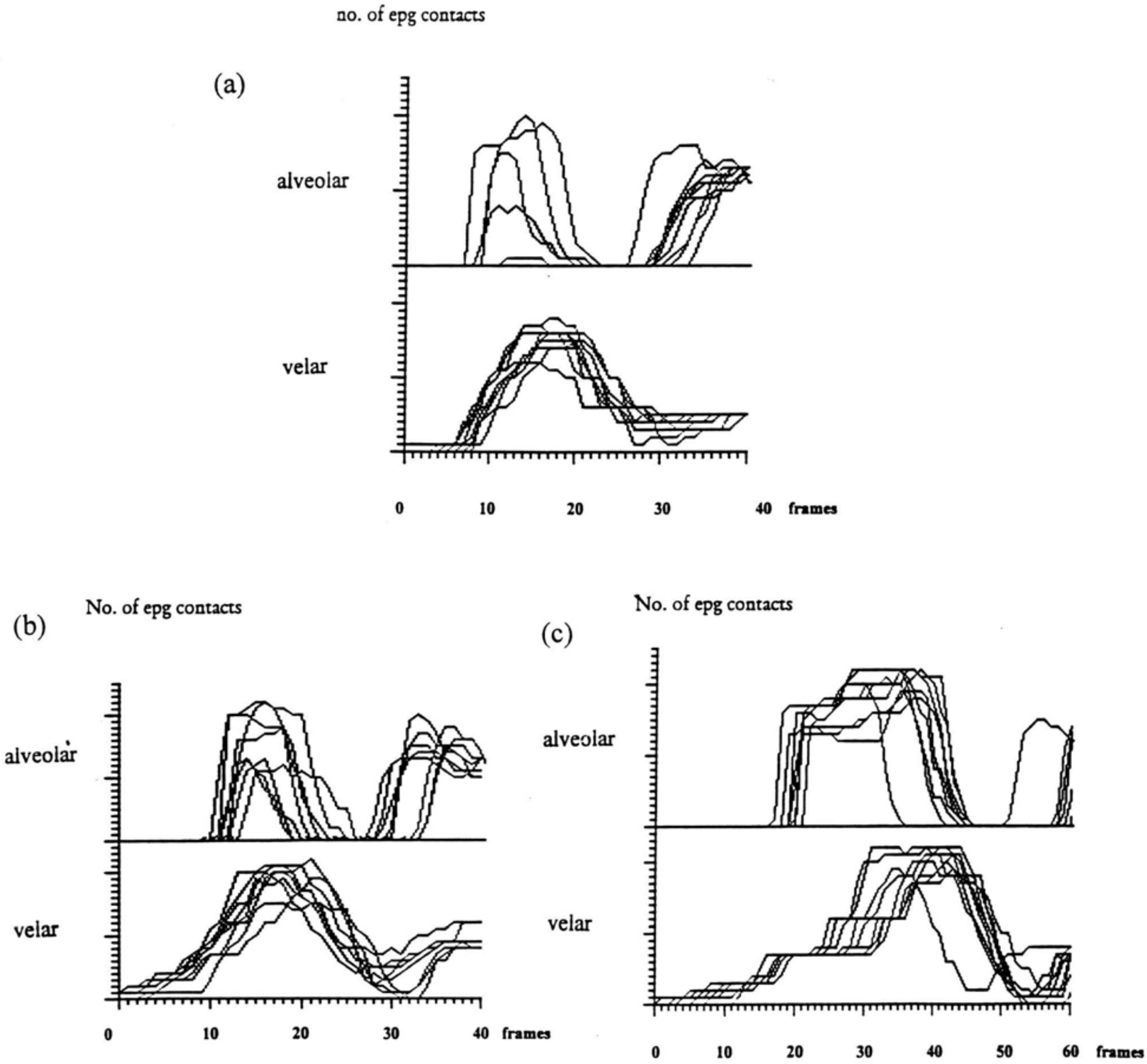


Figure 7 EPG 'totals' graphs showing tongue-palate contact in both alveolar and velar regions for sequence /n#k/ in ...*ban cuts*...The x-axis represents time in frames. 7(a) fast speech, subject JSC; 7(b) fast speech, subject AW; 7(c) careful speech, subject AW

In 7(a), the curves describing residual alveolar articulations are plotted. The shallowest curve of all represents repetition number 5 on Fig. 6 where for 5 EPG frames one electrode was contacted in the third row. The next shallowest curve is the fourth repetition on Fig. 6 where only a partial occlusion is formed and with a slightly more gradual onset than reps 1-3. Of course only 5 'alveolar' curves are shown on this graph because there was no alveolar gesture discernible on the EPG trace for the other 5 repetitions.

AW was one of only two subjects who produced no assimilations for /n#k/ in either careful or fast speech. This subject was quite variable in the fast rate, however, with respect to timing and especially to amount of total contact for /n/ (b). The main reduction strategy adopted was consonant shortening which was sometimes combined with reduction of alveolar contact. This combination produced the variability in alveolar curves in (b), an effect less apparent in (c) for the careful rate. Notice that compared to JSC, the alveolar gesture in (b) is initiated later. Also, the timing relationship between maximum closure for the alveolars and velars for AW in (b) is not the same as that for JSC in (a). AW frequently produced double articulations in careful speech and so maximum contact for each region was often simultaneous. JSC, however, tended to avoid double articulations as illustrated in Fig 6. Kinematics of tongue tip/blade and dorsum are maximally contrasted in 7(c) regarding time interval between initiation of gesture and maximum tongue-palate contact for that gesture. Here for AW velar contact builds up during the markedly less gradual tongue-tip articulation for each repetition. Timing of the release of the alveolar and velar in (c) is likewise consistent apart from one repetition. Of course, in the careful rate (c) the whole sequence up to the complete loss of contact for the /k/ takes around 50 frames whereas for fast speech (a) it takes around 30 frames.

A quite different reduction strategy altogether is used by 3 subjects occupying the right-hand side of the graph in Figure 4 - SW, SM, JR - who all speak Standard Southern British. For all their /n#k/ fast sequences there was a complete absence of alveolar closure and little evidence of lateral contact further forward than that characterising a lexical velar. If lateral contact was in evidence, then for the token to be classified as a residual alveolar according to the definition adopted in this paper, there would be contact along the sides of the palate least one row further forward than for a lexical velar produced by the same speaker under the same speech rate condition. When to compared to the EPG patterns for these subjects' fast all-velar sequences *...bang comes...*, the 'assimilated' /n/ to /k/ patterns are very similar. It would appear that /n/ has been deleted in this context.

For one subject in particular there was even less EPG contact in the velar region for the 'assimilatory' velars than for lexical velars. In other words the habitual posterior tongue placement is more retracted. Some examples are shown in Figure 8 below.

If this was the case, place assimilation here could be accounted for by Articulatory Phonology where gestures can reduce in magnitude and overlap, often to the extent that place of articulation is lost, although segments are never 'deleted'. If so-called complete alveolar assimilation is not necessarily incompatible with tongue tip elevation then the next question is whether the tongue tip is raised for some or all of the habitual complete assimilations illustrated in Fig 8 (a) & (b). But for those residual articulations produced by JSC and WCM where the tongue is making quite advanced contact with the sides of the teeth leaving less area of the tongue-tip /blade to manoeuvre, it might be the case that tongue-tip elevation is somewhat inhibited. But since the tongue-tip can function as a semi-independent articulator in relation to the tongue dorsum, it is possible that the tongue-tip is still able to describe a substantial raising (and looping?) trajectory although not of the same magnitude as that which, it is suggested, may give rise to the type of retracted velar seen in Fig 8 (a) & (b).

4. Conclusion

The two most interesting questions raised from this study have resulted from the identification from this data set of four broad inter-speaker reduction strategies (namely, one which reduces the consonant but avoids assimilation; one which avoids non-assimilation, one which involves binary variation and one which involves non-binary variation). Firstly, why does the distribution of residual alveolar gestures for this data set concentrate around only 2 subjects? While JSC and WCM produce full alveolars and 'complete' assimilations *aswell*, for subjects TH and FG the production of either of these is the result of a categorical binary option, intermediary articulations being somehow 'blocked'. To what extent can this distinction between these two groups of speakers be attributed to the absence or presence of something in their mental representation?

Electropalatography is a technique which can provide useful information about tongue-palate contact during connected speech leading us to the formation of important research questions. It cannot, however, provide information on proximity of the tongue to passive articulators and does not give an indication of which part of the tongue is involved in contact. In order to confirm or challenge the limited interpretations of the EPG data presented here, a combined EMA/EPG experiment using similar stimuli is planned involving a subject from each of the two groups. EMA tracks the movement, in the x-y plane, of sensors attached to the mid-line of the tongue and provides complementary information to EPG on tongue-dynamics and the overall configuration of the tongue during connected speech. The complete assimilations produced by the 'binary' subjects and those produced by the 'non-binary' subjects will hopefully be distinct in terms of tongue-tip elevation. The hypothesis is that the complete assimilations of the former group will involve less or no tongue-tip elevation compared to the latter, since for the latter the EPG evidence suggests that contact with the alveolar ridge is still being targeted. Clear, if reduced coronal trajectories, falling short of contact could be in evidence some of the time or on a more consistent basis. Kühnert (1993) in her study of the assimilatory behaviour of voiceless plosives reported that in cases of EPG-defined complete assimilation, clear coronal elevations are sometimes present and sometimes are not. Speaker-specific reduction strategies as reported here could prove to be the cause of this variability. The second question which could be answered by EMA observations of tongue-tip movement during assimilation, evolves from the type of data shown in Fig 8 (a) & (b).

These speakers are either consistently preserving the raising gesture part of the alveolar target, which could explain the retracted velar position and lack of lateral contact, or they are producing an extreme form of assimilation, or possibly alternating these articulations. The identification of habitual articulations from EPG data may not be mirrored in EMA data for the same sequences.

References

Barry, M. (1985) A palatographic study of connected speech processes. *Cambridge Papers in Phonetics and Experimental Linguistics* 4

Barry, M. (1988) Assimilation in English and Russian. Paper presented at the colloquium of the British Association of Academic Phoneticians, Trinity College, Dublin, March 1988

Browman, C.P. and Goldstein, L.M. (1990) Tiers in articulatory phonology, with some implications for casual speech. In Kingston, J. & Beckman, M.E. (eds.) *Papers in Laboratory Phonology I* Cambridge University Press, 341-76

Brown, G. (1990) *Listening to Spoken English.* Longman, Harlow 2nd edition

Farnetani, E. and Bùsa, M.G. (1994) Italian clusters in continuous speech. *Proceedings of 1994 International Conference on Spoken Language Processing* 1, 359-62 Yokohama, 18-22 September

Gay, T.J. (1981) Mechanisms of the control of speech rate. *Phonetica* 38, 148-58

Hardcastle, W.J. (1994) Assimilation of alveolar stops and nasals in connected speech. In J. Windsor Lewis (ed.) *Studies in General and English Phonetics in Honour of Professor J.D. O'Connor.* Routledge, 49-67

Hardcastle, W.J. (1985) Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4, 247-63

Hardcastle, W.J. and Roach, P.J. (1979) An instrumental investigation of coarticulation in stop consonant sequences. In P. Hollien and H. Hollien (eds.) *Current Issues in the Phonetic Sciences* Amsterdam: John Benjamins A.G. 531-540

Jun, J. (1996) Place assimilation is not the result of gestural overlap: evidence from Korean and English. *Phonology* Cambridge University Press 13, 377-407

Kerswill, S. (1985) A sociophonetic study of connected speech processes in Cambridge English: an outline and some results. *Cambridge papers in Phonetics and experimental Linguistics* 4

Kerswill, S. (1987) Levels of linguistic variation in Durham. *Journal of Linguistics* 23, 23-49

Kühnert, B. (1993) Some kinematic aspects of alveolar to velar assimilations. *Forschungsberichte des Institut fur Phonetic und Sprachliche Kommunikation der Univeritat Munchen* 31, 263-72

Lindblom, B. (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-81

Lindblom, B. (1983) Economy of speech gestures. In P.F. McNeilage (ed.) *The Production of Speech* New York: Springer Verlag 217-45

Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H and H theory. In Hardcastle, W.J. and Marchal, A. (eds.) *Speech Production and Speech Modelling* Dordrecht: Kluwer

Nolan, F. (1992) The descriptive role of segments: evidence from assimilation. In D.R. Ladd and G.J. Docherty (eds.) *Papers in Laboratory Phonology II* Cambridge University Press, 261-80

Wright, S. and Kerswill, P. (1989) Electropalatography in the study of connected speech processes. *Clinical Linguistics and Phonetics* 3, 49-57

AN INSTRUMENTAL STUDY OF ALVEOLAR TO VELAR ASSIMILATION IN CAREFUL AND FAST SPEECH

Lucy Ellis and William J. Hardcastle
Department of Speech and Language Sciences, Queen Margaret University College, Edinburgh, U.K.

ABSTRACT

The assimilation of a word-final alveolar to a following velar has been traditionally described as a discrete phonological process. That is, the place of articulation features for the alveolar have been completely swapped for those of the velar. More recently electropalatographic (EPG) studies have shown empirically that this process is sometimes gradual, providing evidence of intermediate 'residual' alveolar articulations. These conflicting perspectives raise the question: at what level in the generation and execution of an utterance *does* assimilation occur? A number of speakers' productions of /n#k/ were recorded using EPG. The major finding is that while some subjects produce gradient assimilations, others clearly demonstrate categorical assimilations. The lack of residual movement in the latter group was confirmed in a pilot study using EPG in combination with EMA (electromagnetic articulography), a technique which complements EPG contact-only data. On the basis of this, a speaker-specific model of assimilatory behaviour is proposed.

1. EPG STUDY

1.1 Introduction

Assimilation describes the variation in the phonetic description of a speech unit as it becomes more like an adjacent speech unit. For example *red coat* may be realised as [ɹɛg kəʊt]. This process is particularly prevalent in fast speech.

Standard phonological theory traditionally describes place assimilation as a discrete phonological process [1]. So, in the case of alveolar to velar assimilation either there is no assimilation and the alveolar target is preserved or an assimilation takes place where the alveolar target is completely replaced by a velar target. An assimilation is said to be caused by a cognitive substitution rule, utilising stored phonetic/phonological information, which results in categorical variation. More recently, however, the level of phonetics, previously thought of as mere implementation of phonological form determined at a higher cognitive level, has been regarded by some as having considerable explanatory power in accounting for assimilatory processes. Research into the articulatory mechanisms underlying assimilation have shown that this process is sometimes gradual [2,3]. In the case of alveolar to velar assimilation there may be intermediate assimilatory forms which indicate 'residual' alveolars where the tongue's target has been 'undershot'. Thus, phonological theory and instrumental research have tended to locate assimilation within the abstract planning stage and the concrete physical execution stage respectively.

Previous studies in this area were based on small subject numbers (typically one or two) and were concerned more with identifying types of assimilatory pattern than with how speakers might differ with regard to preferred assimilatory strategy. For these reasons the EPG study reported below is an attempt to systematically investigate the distribution of types of alveolar assimilation across a number of speakers. Another way in which work reported in this paper sought to overcome some limitations of previous EPG studies was through the use of Electromagnetic Articulography (EMA) in combination with EPG in a follow-up pilot study. EPG data in isolation gives information on tongue-palate contact only and not on underlying movements. Thus apparently complete alveolar assimilations, indistinguishable from lexical velar patterns, may involve some tongue tip raising short of contact with the alveolar ridge, indicating a reduced but preserved alveolar gesture. This pilot study will be presented in section 2.

In this paper we will show the relevance of speakers' asssimilatory strategy to internal modeling of speech planning.

1.2 Method

1.2.1 Stimuli Speech material captured a potential site of alveolar assimilation and a neutral velar control sequence. These experimental combinations were embedded in the sentences: *"It's hard to believe the ba<u>n</u> <u>c</u>uts no ice "* /n#k/ and *"I've heard the ba<u>ng</u> <u>c</u>omes as a big surprise"* /ŋ#k/. Another experimental sentence was devised to capture the alveolar to alveolar sequence /n#t/: *"I'm not surprised the ba<u>n</u> <u>t</u>ouched a raw nerve"* so that coarticulatory effects on /n/ before /k/ can be compared with /n/ in a non-coarticulatory context. The patterns for the velar to velar control sequences served as a comparison for apparent cases of complete alveolar assimilation and as a yardstick for the identification of residual alveolars. The vocalic environment for the sequences was kept as consistent as possible. Bordering vowels were /a/ & /ʌ/ and the /a/ vowel was preceded by a bilabial stop to eliminate the possibility of any lingual coarticulatory effects on the target consonants. A further 4 filler sentences of no experimental interest were added to the original 3. 10 repetitions of each sentence were required bringing the total stimuli to 70 sentences.

1.2.2 Data collection The technique of electropalatography (Reading EPG3 system) was used to record the timing and the location of tongue contact with the hard palate during continuous speech. 10 speakers with EPG palates were recorded.

The experiment fell into two parts. The aim of the first part was to elicit careful speech and all subjects were instructed to read each sentence *slowly and clearly*. The aim of the second part was to elicit fast and casual speech. The material for this second part was identical to the careful speech part but the 70 sentences were arranged in groups of 3 and filler sentences were purposely distributed to avoid 'clustering' of experimental sentences. 2 filler items were repeated an extra time each to make 24 groups of 3 sentences. A time limit of 5 seconds on the delivery of each group of sentences was imposed on the speakers. The time constraint automatically ruled out any attempt on the part of speakers to pause between each sentence or to impose too complex an intonational structure on the sentences. The time limit was successful in eliciting speech appreciably faster than the careful speech condition and in preventing over-awareness of test items they had already encountered in the first half of the data collection. The sentences in the first half of the experiment and the groups of sentences in the second were individually cued during recording with a pause between each.

1.2.3 Data analysis The criteria for labelling a contact pattern as an alveolar stop closure was the presence of mid-sagittal contact in the first three rows of the EPG palate. Thus according to this definition, Figure 1 (a) is an example of an EPG contact pattern annotated as an alveolar stop articulation, while in 1 (b) we see a contact pattern annotated as an assimilation because there is no mid-sagittal contact in the alveolar region. Individual palate diagrams in 1 (a) and (b) are 10ms apart and tongue-palate contact is indicated by filled circles. The top of each diagram is the alveolar region and the bottom is the velar region. 1 (b) is an example of a residual alveolar articulation whereby closure across the alveolar ridge is absent but the tongue has still made the supporting lateral gesture. A residual alveolar is considered to be an intermediate assimilatory stage where the target closure is undershot due to time constraints. This interpretation follows Lindblom's duration-dependent undershoot model [4]. On the basis of acoustic evidence of vowel reduction he proposed that articulatory and acoustic undershoot of vowels is a function of ***reduction of movement*** towards the vowel target due to physiological limitations.

Figure 1 (a) EPG patterns for /n#k/ fast speech annotated as alveolar stop closure starting at frame 253 (Subject b)

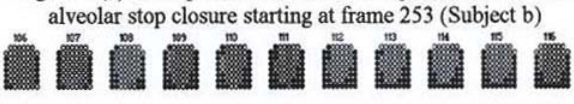


Figure 1 (b) EPG patterns for /n#k/ fast speech annotated as alveolar assimilation (Subject a)

1.3 Results and Discussion

In the fast speech condition a number of speaker specific assimilatory strategies were identified. 2 subjects produced only non-assimilations, 4 subjects always produced what appeared to be complete assimilations and the remaining 4 subjects each produced varying forms of /n#k/. Figure 2 shows the occurrence of assimilations for all speakers. All tokens here are categorized as either non-assimilations or assimilations with residual alveolars/partial assimilations subsumed into the assimilation category, as they were for the annotation procedure. Speakers a-j are ranked from left to right according to frequency of assimilations.

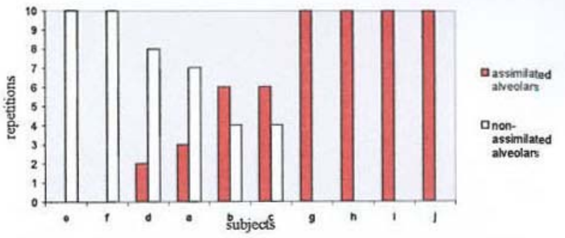


Figure 2 distribution of assimilations and non-assimilations for individual speakers in productions of /n#k/ fast speech

Fast speech /n#k/ EPG patterns for those speakers who produce habitual 'complete' assimilations (g, h, i and j) were indistinguishable from their fast speech lexical /ŋ#k/ productions. It is clear from the graph, however, that assimilation is not motivated by fast speech alone for all speakers (i.e. subjects e and f).

But the most notable result comes from the assimilatory data for those speakers who, in Figure 2, appear to vary between non-assimilation and assimilation. A more detailed examination of the type of patterns produced such as in Figure 1, reveals a fundamental contrast in assimilatory strategy between two groups of subjects, namely Subjects a and b and Subjects c and d.

The contact patterns for subjects c and d suggest the adoption of a binary segmental strategy. That is, either full alveolar contact is achieved for target /n#k/ or an assimilation takes place whereby the place features for /n/ have been completely swapped for those of /ŋ/. The most plausible explanation is that this is achieved for these speakers by accessing a phonological rule. By contrast the EPG data for subjects a and b show a non-binary continuum of assimilatory patterns. Patterns for these speakers can be ranged from full alveolar stop closure through intermediate residual patterns to velar patterns indistinguishable from lexical controls. Figure 3 shows all 10 repetitions of /n#k/ fast speech produced by subject b. Each line captures the tongue-palate contact for a single realization of the target sequence. All 10 realisations are ordered to show an articulatory continuum from full alveolar stop closure at the top to apparently complete assimilation at the bottom. Repetition 3 on line 3 of Figure 1 shows a slightly shorter alveolar articulation with less contact in the alveolar region (rows 1-3) than repetitions 1 and 2. By the time we get to repetition 4, alveolar closure is only partial. Repetition 5 is a residual alveolar articulation whereby the tongue is stretching as far forward as row 3, frames 316-320. The next 5 repetitions follow a progression of lessening side contact until repetition 10, which is the result of either a spirantised /k/ or a closure made too far back for the EPG palate to sample.

In fact, the only 3 residual 'undershoot' articulations

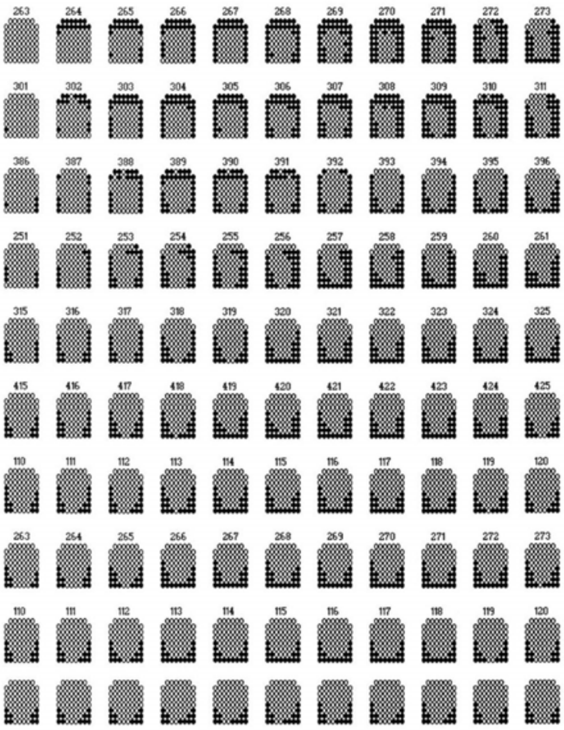


Figure 3 All 10 repetitions of /n#k/ fast speech subject b. Each line captures tongue-palate contact for a single realisation of /n#k/

2. PILOT EPG/EMA STUDY

2.1 Introduction

EPG data is limited because it gives information on tongue-palate contact only and not underlying lingual movement which does not necessarily result in contact with the hard palate. This information is important in the light of speculation that apparently completely assimilated /n#k/ sequences may be accompanied by some residual movement in the form of vertical tongue-tip raising [3]. Kühnert [5] found that identical EPG patterns arising from assimilated alveolar to velar sequences and velar control sequences *may* show different tongue trajectories in the EMA display. This finding undermines the view that differences in lexical form will always result in distinct phonetic output.

If tongue tip raising could be found for the complete assimilations of subjects c or d, then their assimilatory strategy would be gradient and not categorical as previously assumed. This would also mean that gradient movement, on an assimilatory continuum, is observable in both EPG and EMA data since we have already seen gradience with EPG for subjects a and b. Intermediate undershoot forms involving lateral extension of the tongue body visible on EPG patterns as in Figure 1 (b) are surely quite different articulatory events to those forms involving vertical tongue tip raising.

Two subjects who produced categorical assimilations, d and h, were selected to be recorded using combined EPG/EMA to look again for any gradience. Velar to velar sequences /ŋ#k/ were used as controls assuming that these involve minimal or no tongue tip raising.

2.2 Method

Subjects d and h were re-recorded using EPG in combination with EMA (Carstens AG100 Electromagnetic Articulograph). EMA is a transduction device which tracks x-y movement of coils attached to the mid-line of the tongue, typically on the tip/blade, tongue body and tongue dorsum.

Stimuli for this experiment were identical to that of the EPG-only study.

2.3 Results and Discussion

On analysis of the EPG data, it was found that subjects d and h replicated the assimilation strategy they each used in the EPG-only study. On analysis of the EMA data, there was no evidence of gradience in the form of tongue tip displacement for assimilated /n#k/ sequences of a greater magnitude than that for a neutral velar control articulation.

The articulatory positions of tongue tip and tongue dorsum coils for fast speech /n#k/ and /ŋ#k/ tokens at the beginning of the consonant cluster for subject d and h are shown in Figures 4 and 5 below. Beginning of the cluster was defined as the moment of maximum tongue tip displacement (regardless of whether a stop closure was achieved or not) and this was taken from minimum tangential velocity of the coil as it reaches its maximum height and changes direction. x-axis and y-axis position is shown in millimetres. The left hand cluster on each graph shows tongue tip positions and the right hand cluster shows tongue dorsum positions. Subject d's articulatory positions are plotted for non-assimilated alveolar sequences (numbering 5), apparently assimilated alveolar sequences (numbering 5) and all velar control sequences (numbering 10). For subject h, articulatory positions are plotted for apparently assimilated alveolar sequences (10) and velar control sequences (10) only. Subject h produced only 'complete' assimilations.

In Figure 4 maximum vertical displacement for the tongue tip is defined by full alveolar closure for the 5 non-assimilations produced by Subject d. For this speaker, however, it is clear that the tongue tip cluster for the 'complete' /n/ assimilations overlaps with the cluster for the underlying velars. This means that there is no vertical displacement for any assimilated /n#k/ sequence beyond that which accompanies a neutral velar control sequence and thus no intermediate partial assimilation stage in between full alveolar stop closure and complete assimilation. In fact the tongue tip cluster for assimilated /n/ is more constrained spatially than the cluster for underlying /ŋ/ in a way more characteristic of a target articulation. The 'target' in this case could be the result of a high-level instruction to delete the coronal gesture.

For Subject h, a similar picture emerges with overlap of tongue tip positions for assimilated /n/ and lexical /ŋ/. Since this subject produced no full alveolar stop closures for /n#k/ no maximum tongue tip raising is defined, but it seems that the raising movement for underlying velars is surprisingly advanced for some tokens. Once again, the height range for underlying forms is greater than that for assimilated /n/ forms suggesting a constraint on movement variability for the latter. These results confirm that Subject d and h are genuinely operating categorical assimilatory strategies.

could be the result of a high-level instruction to delete the coronal gesture.

For Subject h, a similar picture emerges with overlap of tongue tip positions for assimilated /n/ and lexical /ŋ/. Since this subject produced no full alveolar stop closures for /n#k/ no maximum tongue tip raising is defined, but it seems that the raising movement for underlying velars is surprisingly advanced for some tokens. Once again, the height range for underlying forms is greater than that for assimilated /n/ forms suggesting a constraint on movement variability for the latter. These results confirm that Subject d and h are genuinely operating categorical assimilatory strategies.

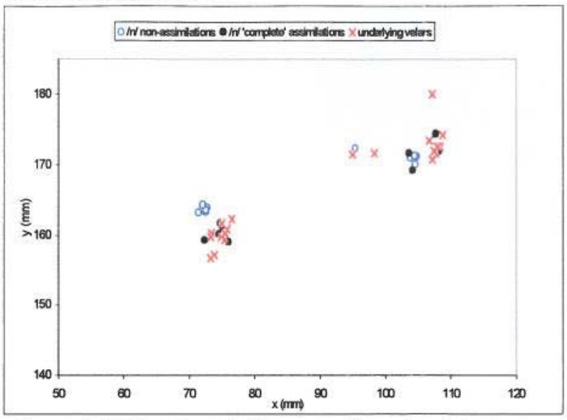


Figure 4 Subject d: articulatory positions (mm) for tongue tip (left cluster) and tongue dorsum (right cluster) at the moment of minimum tangential velocity for all non-assimilated alveolar tokens, assimilated alveolar tokens and all underlying /ŋ/ tokens

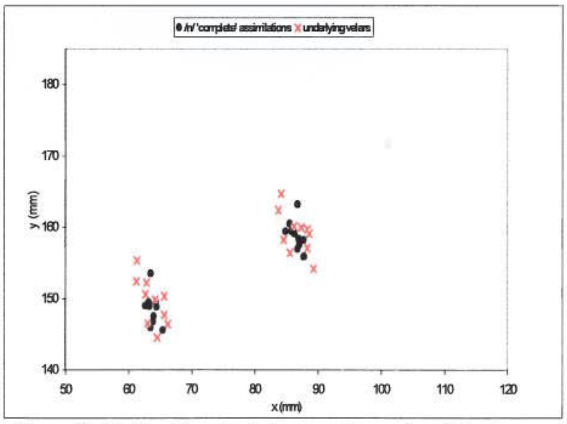


Figure 5 Subject h articulatory positions (mm) for tongue tip and tongue dorsum at the moment of minimum tangential velocity for all assimilated alveolar tokens and all underlying /ŋ/ tokens

3. CONCLUSIONS

There are two principal conclusions to be made from the results of these studies. The first concerns a possible speaker specific basis for assimilatory behaviour and the second addresses the success of the methodologies employed.

The specific contribution of this work to knowledge about assimilatory behaviour is not that speakers have a preferred assimilatory strategy but that these preferred strategies can be fundamentally different. For some speakers assimilation appears to have a mechanical basis, motivated by phonetic factors such as speech rate. For others optional assimilation of this kind seems to be governed by the application of a cognitive rule. We can surmise from this that speaker specific and language specific aspects of speech can override language universal, biomechanical and other constraints on assimilatory processes.

It is clear that EMA used in combination with EPG is a powerful and promising tool in the pursuit of answers to questions about assimilatory strategy. It provides complementary information on the midsagittal and the lateral plane and combines more fine-grained resolution in the anterior region of the vocal tract from EPG with more extensive information on velar movement from EMA. The addition of EMA data to EPG data collected in the pilot study reported above, however, did not result in a radical redefinition of Subject d and h's assimilatory strategy. For this reason the limitations of EPG data may not be as serious as previously assumed.

REFERENCES

[1] Spencer, A. 1996 *Phonology: Theory and Description* Oxford: Blackwell

[2] Hardcastle, W.J. 1994 Assimilation of alveolar stops and nasals in connected speech. In *Studies in General and English Phonetics in honour of Professor J.D. O'Connor* J. Windsor Lewis (editor). Routledge, 49-67

[3] Wright, S. and Kerswill, P. 1989 Electropalatography in the study of connected speech processes. *Clinical Linguistics and Phonetics* 3, 49-57

[4] Lindblom, B. 1963 Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-81

[5] Kuhnert, B. 1993 Some kinematic aspects of alveolar to velar assimilations. *Forschungsberichte des Institut fur Phonetic und Sprachliche Kommunikation der Univeritat Munchen* 31, 263-72